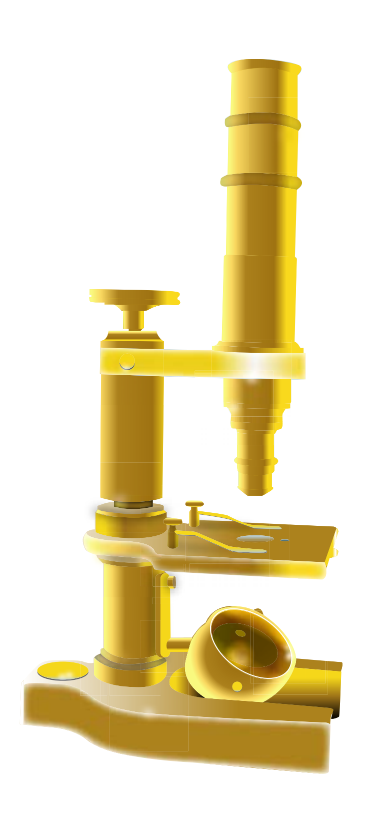
Sciences de gestion

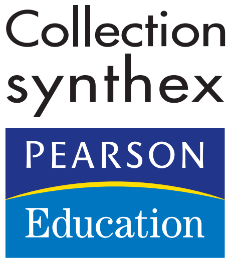
Synthèse de cours & exercices corrigés

Économétrie



- Cours et exercices adaptés aux besoins des économistes et des gestionnaires
- Corrigés détaillés avec Excel, SPSS, TSP, Easyreg
- Données utiles aux exercices sur www.pearson.fr

Collection synthex



Éric DOR

Économétrie

Sciences de gestion

Synthèse de cours



Exercices corrigés

Économétrie

Éric DOR
professeur associé d'Économétrie
à l'IESEG *School of Management* (Lille)

Direction de collection : Roland Gillet
professeur à l'université Paris 1 Panthéon-Sorbonne

Collection
synthex

PEARSON
Education
France

ISBN : 978-2-7440-4071-9

Composition sous LaTeX : ScripTeX



Sommaire

L'auteur

Éric Dor est docteur ès sciences économiques. Il est directeur de la recherche et professeur associé à l'IESEG School of Management de Lille, membre de la Conférence des Grandes Écoles de France. Il enseigne également à l'Institut Catholique des Hautes Études Commerciales (ICHEC) de Bruxelles. Il est l'auteur de nombreuses publications scientifiques, en particulier dans des revues comme *Oxford Bulletin of Economics and Statistics*, *Empirical Economics*, *Recherches Économiques de Louvain*, et *Recherches et Applications en Marketing*. Au cours de sa carrière, il a été Senior Economist chez Wharton Econometric Forecasting Associates. Il a été fréquemment maître de conférences invité à l'Université Catholique de Louvain et a été invité dans plusieurs centres de recherche internationaux, dont le Graduate Center de la City University of New York.

Introduction

L'approche de ce livre est résolument pédagogique. Son objectif est de présenter clairement les principales méthodes économétriques et d'expliquer en détail comment les utiliser en pratique. Notre ouvrage se distingue par l'abondance des études de cas exposées, qui utilisent systématiquement des données réelles et qui portent aussi bien sur des problématiques d'entreprise que sur des problématiques financières ou macroéconomiques. Ce livre constitue donc un outil particulièrement utile à l'apprentissage de l'économétrie par des étudiants en sciences de gestion comme en sciences économiques.

L'ouvrage se distingue également par la place qu'il accorde à expliquer comment les modèles sont spécifiés pour différents types d'applications. L'enseignement de l'économétrie se concentre trop souvent exclusivement sur les techniques d'estimation des modèles, sans détailler au préalable les méthodes de spécification de ces modèles. Or, dans la pratique, la validité d'une étude économétrique dépend de la pertinence de la spécification du modèle estimé ; il est vain de connaître les différentes méthodes d'estimation et d'inférence statistique si on les applique à des modèles incohérents.

Toutes les données utilisées dans les exercices peuvent être téléchargées sur le site Internet de l'éditeur, à l'adresse **www.pearsoneducation.fr**. Les applications sont réalisées à l'aide de différents logiciels, dont l'usage est très répandu. D'une part, pour certains exercices simples, nous montrons comment réaliser des calculs économétriques avec un logiciel de type tableur, Excel, en raison de sa popularité sur les postes de travail. D'autre part, nous initions le lecteur à l'utilisation de logiciels économétriques spécialisés de grande qualité : TSP, SPSS et Easyreg. Ceux-ci sont complémentaires : ils diffèrent dans leur mode de fonctionnement, ce qui donne au lecteur toutes les clés des outils informatiques – TSP est basé sur la programmation de séquences d'instruction tandis que SPSS et Easyreg reposent sur des choix de menus. Pour chacun des logiciels utilisés, le livre présente une introduction détaillée à son utilisation de base. De cette manière, le lecteur peut passer à une mise en pratique immédiatement, sans avoir à lire au préalable les notices d'utilisation fournies par les éditeurs. Toutefois, notre ouvrage ne prétend pas se substituer à la documentation officielle, dont la lecture est indispensable pour une utilisation approfondie. Précisons également que le choix de ces logiciels n'implique pas de jugement de valeur quant aux autres outils économétriques qui existent sur le marché – il n'était pas possible d'inclure une présentation détaillée de tous les logiciels disponibles.

La compréhension de l'ouvrage nécessite la connaissance de quelques notions mathématiques de base. Au besoin, le lecteur peut se référer à l'ouvrage Mathématiques appliquées à la gestion de Ariane Szafarz. De la même manière, une connaissance de base de la théorie statistique est nécessaire. Le lecteur peut se reporter utilement au livre de Patrick Roger : Probabilités, statistique et processus stochastiques, publié dans la même collection. Les méthodes d'estimation et leurs propriétés sont présentées avec une grande rigueur mathématique et statistique, tout en s'efforçant d'expliquer la portée pratique des résultats présentés ; le lecteur doit comprendre sous quelles conditions chaque méthode ou chaque test peut être utilisé à bon escient. Les preuves mathématiques des différents résultats et propriétés ne sont toutefois pas détaillées dans cet ouvrage, le lecteur intéressé étant renvoyé pour cela aux nombreux ouvrages d'économétrie théorique existants. Notre conviction est que l'enseignement de l'économétrie doit d'abord intéresser l'étudiant à la discipline en lui montrant d'emblée les applications pratiques enthousiasmantes qu'elle permet de réaliser. La motivation qui en résulte devrait inciter naturellement le lecteur à approfondir ensuite sa connaissance de l'économétrie, en s'intéressant aux développements mathématiques à la source des méthodes et de leurs propriétés.

Cet ouvrage constitue le manuel idéal pour un premier cours d'économétrie, centré sur l'explication des méthodes et sur leur mise en pratique. Le professeur peut y ajouter lui-même, à sa propre convenance, les démonstrations mathématiques de certains résultats. Dans les programmes d'enseignement où l'on organise séparément des cours d'économétrie théorique et un cours d'économétrie appliquée, notre ouvrage constitue bien sûr un manuel approprié à ce dernier. Ce livre peut également être utilisé en complément d'un manuel essentiellement théorique.

Je tiens à remercier TSP International pour m'avoir autorisé à reproduire ici des extraits de résultats produits avec le logiciel TSP, Herman Bierens pour avoir permis la reproduction de captures d'écran issues d'Easyreg, et SPSS France pour un accord similaire concernant SPSS. La reproduction d'éléments issus d'Excel respecte les conditions imposées par Microsoft Corporation, telles qu'elles étaient publiées sur la page Internet **http ://www.microsoft.com/france/permission/copyrgt/cop-img.htm#ScreenShot** en février 2004.

Je remercie également Roland Gillet, le directeur de la collection, pour la confiance qu'il m'a témoignée en me proposant de rédiger ce manuel, ainsi que Pearson Education France pour le soin apporté à la réalisation de l'ouvrage, en particulier Pascale Pernet, Antoine Chéret, et tout spécialement Christophe Lenne pour son engagement, sa patience et sa rigueur. Leur professionnalisme permet de proposer au lecteur un produit de grande qualité.

Éric Dor
Docteur ès sciences économiques
Directeur de la recherche
IESEG School of Management
Lille



Modélisation en économie et gestion

Ce chapitre définit l'objectif et la méthode générale de l'économétrie. Il précise quelques notions de base indispensables à la compréhension de l'ouvrage, liées à la modélisation mathématique des phénomènes rencontrés en sciences économiques et en sciences de gestion.

1 Utilité et définition de l'économétrie

L'économétrie est le principal outil d'analyse quantitative utilisé par les économistes et gestionnaires dans divers domaines d'application, comme la macroéconomie, la finance ou le marketing. Les méthodes de l'économétrie permettent de vérifier l'existence de certaines relations entre des phénomènes économiques, et de mesurer concrètement ces relations, sur la base d'observations de faits réels.

Dans son acception la plus restreinte, l'économétrie est un ensemble de techniques utilisant la statistique mathématique qui vérifient la validité empirique des relations supposées entre les phénomènes économiques et mesurent les paramètres de ces relations. Au sens large, l'économétrie est l'art de construire et d'estimer des modèles empiriques adéquats par rapport aux caractéristiques de la réalité, et intelligibles au regard de la théorie économique.

2 Relations économiques

La réflexion que l'on peut mener sur une réalité économique quelconque conduit toujours à établir des relations entre les phénomènes économiques concernés. Une réflexion approfondie dans un domaine de science économique ou science de gestion est à la base de toute analyse économétrique. En d'autres termes, la réalisation de travaux économétriques suppose la connaissance préalable des disciplines économiques en jeu, puisqu'elles suggèrent le type de relation à vérifier sur les données réelles observées.

Exemple

On suppose que la consommation totale des ménages augmente avec leur revenu disponible réel, mais diminue quand le taux d'intérêt monte. Une telle relation économique s'écrit de la manière suivante :

$$c = f(yd, r), \quad \text{avec } \frac{\partial c}{\partial yd} > 0 \text{ et } \frac{fc}{fr} < 0 \qquad \text{(a)}$$

où c correspond à la consommation, yd au revenu disponible et r au taux d'intérêt. La notation $f(,)$ désigne une fonction quelconque, linéaire ou non (il faudrait poser des hypothèses supplémentaires pour en préciser la forme fonctionnelle, mais ce n'est pas le propos de cette section). La supposition de départ se formule de la façon suivante : la dérivée partielle de f par rapport à yd est positive – à taux d'intérêt r inchangé, une augmentation du revenu disponible yd implique une augmentation de la consommation c – et la dérivée partielle de f par rapport à r est négative – à revenu disponible inchangé, une augmentation du taux d'intérêt r implique une diminution de la consommation c.

Exemple

Une relation économique suggère que le taux d'intérêt nominal R est une fonction croissante du taux d'inflation INF et du taux de croissance de la production CR :

$$R = f(INF, CR), \quad \text{avec } \frac{\partial R}{\partial INF} > 0 \text{ et } \frac{\partial R}{\partial CR} > 0 \qquad \text{(b)}$$

3 Vérification de l'adéquation empirique des relations

Pour expliquer comment se détermine(nt) un ou plusieurs phénomènes économiques, on construit un modèle à partir de certaines hypothèses et des résultats qu'elles donnent dans le cadre d'une théorie particulière. On vérifie que ce modèle décrit réellement la manière dont le ou les concept(s) d'intérêt se détermine(nt) dans la réalité. Il faut pour cela disposer de mesures réelles des phénomènes (les « statistiques ») et vérifier au moyen de techniques issues de la statistique mathématique (1) que le modèle correspond à ces données observées.

1. Si nécessaire, quelques rappels utiles de la statistique mathématique peuvent être puisés dans tout bon manuel de base, comme par exemple le livre de Probabilités, statistique et processus stochastiques de Patrick Roger, publié chez Pearson Education France dans la même collection.

4 Mesure des taux de réaction

Dans la mesure où le modèle est acceptable, on souhaite également mesurer quantitativement les taux de réaction des phénomènes expliqués aux variations des phénomènes explicatifs. Ces mesures permettront de simuler ultérieurement l'effet de telle ou telle variation hypothétique d'un phénomène explicatif sur les phénomènes expliqués.

Soit un modèle explicatif du taux d'intérêt, sous la forme d'une équation où le taux d'inflation est une variable explicative. On vérifie son adéquation à la réalité observée. Comme on dispose alors des mesures des taux de réaction du taux d'intérêt à ses déterminants, on peut évaluer à l'avance l'effet sur le taux d'intérêt d'une accélération de l'inflation d'un montant déterminé. Souvent, plusieurs théories concurrentes expliquent les mêmes réalités économiques. Les techniques économétriques permettent d'identifier celle qui explique le mieux la réalité, celle qui est au plus près des observations.

5 Formes fonctionnelles et paramètres

L'objectif est de vérifier l'adéquation d'un modèle à la réalité observée et de mesurer les taux de réaction des phénomènes expliqués aux phénomènes explicatifs. Pour confronter efficacement modèle et données, il convient d'exprimer ce dernier sous une forme « manipulable ». Selon la relation (a), la consommation est une fonction du revenu disponible et du taux d'intérêt. Cette formulation est mathématiquement trop « vague » pour pouvoir être confrontée à la réalité observée. Pour pallier le problème, il faut spécifier *a priori* une forme fonctionnelle particulière de la fonction $f()$. Les possibilités sont innombrables.

5.1 Choix d'une relation linéaire

Le choix le plus simple est celui d'une relation linéaire. Il se justifie quand on peut raisonnablement supposer que les dérivées partielles de la variable dépendante par rapport à chaque variable explicative ne sont pas fonction des niveaux atteints par ces variables explicatives. Cette hypothèse signifie que la variation de la variable dépendante, suite à une variation de une unité de l'une des variables explicatives, est toujours la même quels que soient les niveaux déjà atteints par celles-ci.

Exemple

On suppose que la fonction $f()$ est linéaire. Soient les paramètres α, β et γ tels que :

$$c = \alpha + \beta yd + \gamma r \,, \quad \text{avec} \;\; \beta > 0 \;\; \text{et} \;\; \gamma < 0 \qquad (a')$$

On a donc $f(yd, r) = \alpha + \beta yd + \gamma r$. On remarque que :

$$\beta = \frac{\partial c}{\partial yd} \quad \text{et} \quad \gamma = \frac{\partial c}{\partial r}$$

Le coefficient β est donc la dérivée partielle de c par rapport à yd. Il rend compte de l'importance de la variation de c quand yd augmente de une unité, à r constant. Que se passe-il quand r ne change pas, mais que yd augmente de une unité (il s'agit de l'unité dans laquelle yd est exprimé) ? La réponse est que c varie de β unités (il s'agit ici de l'unité de mesure dans laquelle c est exprimé). De la même manière, γ est la dérivée partielle de c par rapport à r. Il rend compte de l'importance de la variation (par exemple en milliards d'euros à prix constants) de c quand r augmente de une unité (par exemple d'un montant absolu de 1 % lorsque r est exprimé en pourcentage), yd restant inchangé. Lorsque la relation entre les variables est supposée linéaire, chaque paramètre est interprété comme la dérivée partielle de la variable dépendante par

rapport à la variable explicative concernée. Chaque paramètre mesure donc la variation de la variable dépendante suite à une augmentation de une unité de la variable explicative concernée, les autres variables explicatives restant inchangées.

5.2 Choix d'une relation non linéaire

La linéarité est certes commode, mais n'est pas toujours une propriété adéquate à la relation traitée. Souvent, il est irréaliste de supposer que la variation de la variable dépendante est toujours la même, suite à une variation de une unité d'une variable explicative, quels que soient les niveaux déjà atteints par cette dernière et par les autres variables explicatives. On ne peut alors partir du principe que les dérivées partielles sont indépendantes des niveaux des variables. Dans ce cas, on travaille avec des relations formalisées sous la forme d'équations non linéaires.

Exemple

On souhaite modéliser la relation entre les ventes d'un produit de grande consommation V et les dépenses de publicité PUB de l'entreprise productrice. Si l'on pense que la « productivité », en termes de ventes, des dépenses de publicité décroît avec leur montant, on peut écrire :

$$V = \alpha \text{PUB}^{\beta}, \quad \text{avec} \ \ 0 < \beta < 1$$

Cette spécification implique en effet une dérivée première de V par rapport à PUB, qui décroît avec le montant de PUB. Autrement dit, au fur et à mesure que les dépenses publicitaires augmentent, l'augmentation des ventes devient de plus en plus faible.

Certaines relations non linéaires sont équivalentes à des relations linéaires entre des transformations des variables.

Exemple

Si l'on transforme les variables en logarithmes, une fonction de production de Cobb-Douglas, du type $Y = AK^{\beta}L^{\gamma}$, où Y, L et K sont la production, le travail et le capital, implique une relation linéaire entre les **transformations** des variables :

$\ln(Y) = \ln(AK^{\beta}L^{\gamma})$ et donc $\ln(Y) = \ln(A) + \beta \ln(K) + \gamma \ln(L)$. Elle n'implique pas toutefois la constance des productivités marginales, qui restent bien sûr $\frac{\partial Y}{\partial L} = A\gamma K^{\beta}L^{\gamma-1}$ et $\frac{\partial Y}{\partial K} = A\beta K^{\beta-1}L^{\gamma}$. Cette nouvelle équation ne constitue qu'une autre manière d'exprimer la même fonction de production : chacune des deux écritures implique l'autre et les propriétés économiques sont exactement les mêmes. L'écriture en logarithme met en évidence que $\beta = \frac{\partial \ln Y}{\partial \ln K} = \frac{\partial Y}{\partial K}\frac{K}{Y}$ et $\gamma = \frac{\partial \ln Y}{\partial \ln L} = \frac{\partial Y}{\partial L}\frac{L}{Y}$ sont les élasticités[1] de la production aux quantités de facteurs capital et travail. Ces élasticités sont supposées constantes (indépendantes des quantités de facteurs K et L) dans une telle fonction de production (Cobb-Douglas). Alors que la dérivée partielle d'une variable x_1 par rapport à une variable x_2 mesure la variation de x_1 (en nombres d'unités) quand x_2 augmente de une unité, l'élasticité de x_1 à x_2 mesure la variation de x_1 (en pourcentage) quand x_2 augmente de 1 %. Les coefficients β et γ, qui ne sont donc pas des productivités marginales, sont des rapports entre productivités marginales et moyennes. La fonction de Cobb-Douglas implique en effet la constance de ces rapports, au sens de leur indépendance par rapport à K et L.

1. Affirmer que l'élasticité de x_1 à x_2 est égale à 2 revient à affirmer la proposition suivante : lorsque x_1 augmente de 1 %, alors x_2 augmente de 2 %.



Cela dit, de nombreuses formes fonctionnelles non linéaires ne peuvent être linéarisées moyennant une transformation des variables.

Exemple

Soit la fonction de production CES, ayant la forme :

$$Y = \left(\lambda K^{-\rho} + (1-\lambda)L^{-\rho}\right)^{-1/\rho},$$

Elle ne peut être linéarisée exactement (c'est-à-dire transformée en une relation linéaire reliant des transformations non linéaires séparées de chaque variable).

Remarque

Une erreur de spécification à éviter : la redondance

Il est important de comprendre l'interprétation des coefficients en termes de dérivées partielles pour éviter des erreurs dans la spécification d'une relation. Une erreur très répandue consiste à introduire une variable explicative supplémentaire sous prétexte qu'elle affecte la variable dépendante par son effet sur une autre variable explicative déjà introduite. C'est le phénomène de la redondance !

6 Validation empirique et types de données

6.1 Dimension du temps ou des agents

Une fois représentées par des formes fonctionnelles adéquates, les relations théoriques, c'est-à-dire le modèle, peuvent être confrontées aux données observées. Il s'agit de vérifier leur caractère explicatif de la réalité et de mesurer concrètement la valeur de leurs paramètres. Il est alors possible de calculer les taux de réaction des variables expliquées aux variables explicatives. Les données observées peuvent être des séries temporelles, des données en coupe instantanée ou des données panel.

Séries temporelles

Quand une équation semble décrire correctement la manière dont une variable évolue d'une période à l'autre, en fonction de l'évolution temporelle de certaines variables explicatives, elle peut être vue comme une relation stable et valable à tout moment. Ses coefficients ne sont pas indicés par le temps. On les suppose constants dans le temps. C'est une hypothèse forte, mais dans la mesure où la théorie économique a une quelconque validité pour expliquer les phénomènes économiques, on peut supposer l'existence de relations stables. Pour les vérifier empiriquement, il faut estimer leurs coefficients à partir des observations historiques des variables du modèle, appelées « séries temporelles » (ou « séries chronologiques »).

Données en coupe instantanée

Quand une équation semble plutôt décrire la manière dont différents agents économiques (entreprises, individus, régions, pays, secteurs...) déterminent la valeur particulière d'une variable en fonction des valeurs que prennent pour eux certaines variables explicatives, elle peut être vue comme une relation commune aux différents agents. Les coefficients

sont supposés les mêmes pour tous les agents, durant une période d'étude donnée. Pour vérifier cette relation, il faut la confronter à des observations concrètes des variables du modèle pour un ensemble d'agents différents, durant une même période. On appelle de telles observations des « données en coupe instantanée ».

Données panel

Quand une équation semble décrire la manière dont une variable varie d'une période à l'autre et diffère d'un agent à l'autre en fonction de l'évolution dans le temps de certaines variables explicatives et de leurs différences d'un agent à l'autre, elle peut être vue comme une relation stable et commune à tous, décrivant le comportement de tous les agents durant toutes les périodes. Pour mesurer et vérifier une telle relation, il faut la confronter à des observations des variables du modèle pour un ensemble d'agents différents, sur des périodes différentes. On appelle de telles observations des « données panel ».

Données réelles ou nominales

Une variable de flux ou de stock peut généralement être mesurée en termes nominaux (à prix courants, en valeur...) ou en termes réels (à prix constants, en volume...). La mesure en termes réels est égale à la mesure en termes nominaux divisée par un indice de prix approprié. Le choix d'un type de mesure au détriment de l'autre dépend logiquement du contexte de la relation étudiée. De manière générale, la variable dépendante et certaines variables explicatives doivent être exprimées en termes réels si la valeur réelle de la variable dépendante reste inchangée quand les valeurs nominales de ces variables explicatives doublent et que tous les prix doublent simultanément.

Certaines variables de taux existent en version nominale ou réelle. C'est le cas des taux d'intérêt et des taux de change. On réalise une approximation du taux d'intérêt réel en calculant la différence entre le taux d'intérêt nominal et le taux d'inflation. On obtient le taux de change réel entre deux devises en multipliant le taux de change nominal par le rapport entre les indices de prix des deux zones concernées. Une fois de plus, le choix de l'une des deux versions est dicté logiquement par le contexte de la relation étudiée.

7 Formulation statistique des relations économiques

En économétrie, on suppose généralement que les variables économiques sont aléatoires. En d'autres termes, on considère que la valeur observée d'un phénomène économique, par exemple l'investissement total effectué durant une année particulière, est en partie due au hasard : c'est la réalisation d'une variable aléatoire correspondante susceptible de produire d'autres réalisations si l'on répéte l'expérience.

Exemple

À chaque période t, on observe la valeur de la variable aléatoire c_t, en l'occurrence la consommation, mais, d'un point de vue conceptuel, on pourrait observer d'autres valeurs, éventuellement différentes, si l'on répétait l'expérience. De la même manière, à chaque période t, les valeurs effectivement observées de yd_t et r_t sont perçues comme des réalisations uniques des variables aléatoires correspondantes yd_t et r_t, qui pourraient avoir d'autres réalisations.

Le hasard détermine en partie les réalisations effectivement observées des variables économiques et les résultats auraient pu être différents. Les probabilités d'obtenir telle ou telle valeur effectivement réalisée sont déterminées par les distributions statistiques des variables. Les relations économiques supposées par la théorie économique imposent des liaisons entre ces distributions.

Exemple

Une relation comme (α′) relie les réalisations particulières des variables aléatoires c_t, yd_t et r_t qu'elle contient, par une forme fonctionnelle avec des coefficients α, β et γ supposés non aléatoires. Habituellement, on ajoute un aléa u_t à la relation :

$$c_t = \alpha + \beta yd_t + \gamma r_t + u_t \qquad (\alpha')$$

(L'exemple est présenté dans un cadre temporel, mais il en va de même en coupe instantanée : $c_i = \alpha + \beta yd_i + \gamma r_i + u_i$, ou en panel : $c_{it} = \alpha + \beta yd_{it} + \gamma r_{it} + u_{it}$.)

On justifie de différentes façons la présence d'un aléa dans une relation entre des variables. Très souvent, on affirme qu'une relation économique n'est pas une représentation exacte et complète de la réalité. Elle ne reprend que les principaux facteurs qui influencent c ; l'aléa u, communément appelé « terme d'erreur », représente tous les effets qui ont tendance à se compenser mutuellement, de toutes les autres variables qui influencent également c. Cette interprétation, très intuitive, a longtemps été favorisée dans les manuels d'économétrie, au détriment des autres, sans que ce choix soit réellement justifié. Selon une autre interprétation (qui n'exclut pas la précédente), très ancienne également, les mesures concrètes des réalisations des variables, telles qu'elles sont calculées et publiées par les instituts de statistiques, s'accompagnent d'erreurs aléatoires et l'aléa u représente l'effet cumulé de toutes ces erreurs sur la relation originale (pour que cette dernière soit exacte, il faudrait que les concepts soient « parfaitement » mesurés). Autre interprétation : si l'on considère que la formule $\alpha + \beta yd_t + \gamma r_t$ constitue une approximation de la variable aléatoire c_t par une fonction des variables aléatoires yd_t et r_t, u_t est l'erreur d'approximation qui en résulte.

8 Processus stochastiques

En finance, en marketing et en macroéconomie, la plupart des données se présentent sous la forme de séries temporelles. Rappelons qu'une série temporelle est un ensemble d'observations qui portent toutes sur un même concept, mais à des dates successives. On suppose qu'à chaque période, la donnée observée est une réalisation (unique) d'une variable aléatoire spécifique, et que l'on obtiendrait d'autres réalisations si l'on répétait l'expérience. On mesure donc la réalisation d'une variable aléatoire (univariée) par période et l'ensemble des variables aléatoires considérées sur les périodes successives forme un processus stochastique. Une série temporellee est une réalisation d'un processus stochastique, au sens où chaque donnée de la série est la réalisation de l'une des variables aléatoires qui composent le processus stochastique.

Les processus stochastiques se répartissent en deux groupes selon qu'ils sont stationnaires ou non. Lorsqu'ils le sont, l'espérance (valeur moyenne) et la variance (dispersion) restent constantes dans le temps, et les covariances entre des composantes de dates différentes

ne dépendent que de l'écart de temps qui les sépare. Un cas particulier de processus stochastique stationnaire est le processus « bruit blanc » (traduction littérale de « white noise ») : l'espérance est nulle en toute période, la variance est constante dans le temps et les covariances entre composantes de dates différentes sont toujours nulles.

Les processus stochastiques non stationnaires se répartissent eux-mêmes en deux groupes selon qu'ils sont à tendance uniquement déterministe ou à tendance stochastique (on les appelle alors « processus intégrés » ou « processus à racine unitaire »). Lorsqu'ils sont à tendance uniquement déterministe, leur non-stationnarité est due à un phénomène purement mécanique ; elle est inhérente à leur partie déterministe, mais en rien à leur partie aléatoire. Lorsqu'ils sont à tendance stochastique, leur non-stationnarité est due à une accumulation progressive de chocs aléatoires ; elle est donc au moins partiellement inhérente à leur partie aléatoire. Ces processus peuvent avoir également une tendance déterministe.

Si l'on travaille avec des séries temporelles, le choix des méthodes d'inférence statistique à employer dépend de la nature des processus stochastiques qui ont généré les données. C'est pourquoi les distinctions évoquées précédemment sont très importantes.

Dans un processus stochastique stationnaire, les coefficients de corrélations entre deux composantes de dates différentes sont appelés « coefficients d'autocorrélation ». Ils ne dépendent que de l'écart de temps, ou retard, qui sépare les deux composantes. La succession de ces coefficients d'autocorrélation, pour des retards croissants, forment ce que l'on appelle un « autocorrélogramme ». Il montre avec quelle intensité les réalisations du processus restent liées linéairement à leurs valeurs passées, pour des retards de plus en plus éloignés.

9 Modèles statiques ou dynamiques et théorie économique

Un modèle statique implique que l'influence d'une variation d'une variable explicative sur la variable dépendante produit tous ses effets durant la période où cette variation a lieu. Il exclut toute inertie et tout délai dans les ajustements de la variable dépendante aux fluctuations des variables explicatives, alors qu'ils sont l'un et l'autre plus la règle que l'exception. En effet, une variable dépendante dépend souvent des valeurs passées, et pas seulement des valeurs actuelles, de ses variables explicatives (délais d'ajustement), ainsi que de sa propre valeur passée (inertie, effets d'habitude). De nombreux phénomènes économiques réels sont donc mieux expliqués par un modèle dynamique plutôt que statique.

Les relations entre les variables que la théorie économique propose sont souvent formulées de manière statique et représentent une situation d'équilibre (plus aucune force économique ne pousse à changer de situation ; tous les ajustements sont effectués). Pour autant, la théorie économique ne prétend pas que, dans la réalité, la situation soit équilibrée à chaque instant. Les données observées rendent compte obligatoirement de cet état de fait.

Il est donc erroné de vérifier une théorie en estimant le modèle statique issu de cette théorie à partir des données observées, car la relation d'équilibre théorique n'est pas vraie à chaque période. Il faut en fait estimer, sur la base de ces données, un modèle dynamique suffisamment riche pour prendre en compte toutes les inerties et délais d'ajustement, et vérifier que la relation entre les variables mises en jeu pour une situation d'équilibre est compatible avec la relation d'équilibre théorique. Pour qu'il en soit ainsi, on peut imposer aux paramètres du modèle dynamique général les contraintes ou restrictions nécessaires.



Un modèle dynamique général relie une variable dépendante à ses valeurs passées et aux valeurs présentes et passées de ses variables explicatives. Ce modèle décrit donc la trajectoire de la variable dépendante en fonction de la trajectoire de ses variables explicatives.

Exemple

Soit une variable dépendante $\ln(Y)$, et ses variables explicatives $\ln(X)$, $\ln(W)$ et $\ln(L)$. Si l'on ne prend qu'une valeur passée pour chaque variable, le modèle dynamique s'écrit comme suit, les deux formes étant équivalentes :

$$\begin{aligned}
\ln Y_t &= \beta_1 + \beta_2 \ln X_t + \beta_3 \ln X_{t-1} + \beta_4 \ln W_t + \beta_5 \ln W_{t-1} \\
&\quad + \beta_6 \ln L_t + \beta_7 \ln L_{t-1} + \beta_8 \ln Y_{t-1} + u_t \\
(\ln Y_t - \ln Y_{t-1}) &= \beta_1 + \beta_2 (\ln X_t - \ln X_{t-1}) + (\beta_2 + \beta_3) \ln X_{t-1} + \beta_4 (\ln W_t - \ln W_{t-1}) \\
&\quad + (\beta_4 + \beta_5) \ln W_{t-1} + \beta_6 (\ln L_t - \ln L_{t-1}) + (\beta_6 + \beta_7) \ln L_{t-1} \\
&\quad + (\beta_8 - 1) \ln Y_{t-1} + u_t
\end{aligned}$$

La solution d'équilibre stationnaire de ce modèle dynamique général est la relation entre les variables qui prévaut dans une situation où elles restent toutes constantes à chaque période, tout en respectant la relation décrite par le modèle dynamique général.

Exemple (suite)

Dans l'exemple précédent, la solution d'équilibre stationnaire est :

$$\ln \tilde{Y} = \beta_1 + \beta_2 \ln \tilde{X} + \beta_3 \ln \tilde{X} + \beta_4 \ln \tilde{W} + \beta_5 \ln \tilde{W} + \beta_6 \tilde{L} + \beta_7 \tilde{L} + \beta_8 \ln \tilde{Y}$$

ou encore :

$$\ln \tilde{Y} = \frac{(\beta_1) + (\beta_2 + \beta_3) \ln \tilde{X} + (\beta_4 + \beta_5) \ln \tilde{W} + (\beta_6 + \beta_7) \ln \tilde{L}}{1 - \beta_8}$$

La solution de croissance équilibrée d'un modèle dynamique général est la relation entre les variables qui prévaut dans une situation où elles croissent au même taux, tout en respectant la relation décrite par le modèle dynamique général.

Exemple (suite)

Dans l'exemple précédent, il faut donc imposer que les taux de croissance de Y, X, W et L, qui sont respectivement donnés par $\ln(Y_t) - \ln(Y_{t-1})$, $\ln(X_t) - \ln(X_{t-1})$, $\ln(W_t) - \ln(W_{t-1})$ et $\ln(L_t) - \ln(L_{t-1})$, soient des constantes :

$$\begin{aligned}
\ln(Y_t) - \ln(Y_{t-1}) &= g_C\,, \quad \forall t \\
\ln(X_t) - \ln(X_{t-1}) &= g_Y\,, \quad \forall t \\
\ln(W_t) - \ln(W_{t-1}) &= g_W\,, \quad \forall t \\
\ln(L_t) - \ln(L_{t-1}) &= g_L\,, \quad \forall t
\end{aligned}$$

La solution de croissance équilibrée est alors :

$$\begin{aligned}
g_Y = \beta_1 + \beta_2 g_X + (\beta_2 + \beta_3) \ln X_{t-1} + \beta_4 g_W + (\beta_4 + \beta_5) \ln W_{t-1} \\
+ \beta_6 g_L + (\beta_6 + \beta_7)\, L_{t-1} + (\beta_8 - 1) \ln Y_{t-1}\,, \quad \forall t
\end{aligned}$$

ou encore :

$$\ln Y_t = \frac{\big(\beta_1 + \beta_2 g_X + \beta_4 g_W + \beta_6 g_L - g_Y\big) + (\beta_2 + \beta_3) \ln X_t + (\beta_4 + \beta_5) \ln W_t + (\beta_6 + \beta_7) \ln L_t}{1 - \beta_8}\,, \quad \forall t$$

Remarque

Parfois, même en croissance équilibrée, certaines variables ne peuvent avoir logiquement qu'une croissance nulle (comme en équilibre stationnaire). C'est généralement le cas des variables de taux, comme les taux d'intérêt.

Exemple (suite)

Si c'est le cas de L dans l'exemple précédent, alors $g_L = 0$ et

$$\ln Y_t = \frac{\left(\beta_1 + \beta_2 g_X + \beta_4 g_W - g_Y\right) + (\beta_2 + \beta_3)\ln X_t + (\beta_4 + \beta_5)\ln W_t + (\beta_6 + \beta_7)\ln \tilde{L}}{1 - \beta_8}, \quad \forall t$$

Parfois la théorie économique suggère qu'à long terme, la variable dépendante doit être proportionnelle à une variable explicative, c'est-à-dire avoir une élasticité unitaire par rapport à cette variable explicative. Il est alors aisé d'identifier les conditions nécessaires sur les coefficients du modèle dynamique général pour que ses solutions d'équilibre soient compatibles avec la théorie. Le mécanisme à correction d'erreur est le modèle qu'on obtient en imposant ces restrictions au modèle linéaire général.

Exemple (suite)

Soit le cas de figure suivant : selon la théorie économique, à long terme Y doit être proportionnel à X, et donc l'élasticité de long terme de Y à X doit être égale à 1. Pour que les solutions d'équilibre du modèle linéaire général soient compatibles avec cette théorie, il faut que $\beta_2 + \beta_3 = 1 - \beta_8$. Le **mécanisme à correction d'erreur** est alors le modèle qu'on obtient en imposant cette restriction au modèle linéaire général :

$$\begin{aligned}(\ln Y_t - \ln Y_{t-1}) = {} & \beta_1 + \beta_2 (\ln X_t - \ln X_{t-1}) + \beta_4 (\ln W_t - \ln W_{t-1}) + (\beta_4 + \beta_5)\ln W_{t-1} \\ & + \beta_6 (\ln L_t - \ln L_{t-1}) + (\beta_6 + \beta_7)\ln L_{t-1} + (\beta_8 - 1)(\ln Y_{t-1} - \ln X_{t-1}) + u_t\end{aligned}$$

Résumé

L'économétrie permet de vérifier l'existence de relations de dépendance entre des phénomènes et de mesurer les taux de réaction qui caractérisent ces relations, en utilisant des données observées. Pour réaliser ces objectifs, toute relation doit d'abord être exprimée mathématiquement au moyen d'une forme fonctionnelle appropriée. La linéarité ne se justifie que lorsqu'il est réaliste de supposer que l'impact d'une même variation d'une variable explicative sur la variable dépendante est toujours le même, quels que soient les niveaux des variables au départ. L'économétrie reconnaît d'emblée le caractère stochastique des phénomènes qu'elle étudie. Les variables observées sont ainsi considérées comme des réalisations de variables aléatoires et les modèles spécifiés sont perçus comme pertinents à un terme d'erreur aléatoire près. Il est souvent nécessaire de recourir à des modèles dynamiques pour rendre compte de l'inertie des comportements. Pour plus de détails sur la nature de l'économétrie et sur certains points développés dans ce chapitre, on peut se référer à Johnston et DiNardo [JOH 1997], Hendry [HEN 1995] et Spanos [SPA 1986].



Problèmes et exercices

EXERCICE 1 VENTES ET PUBLICITÉ

Énoncé

Les ventes V d'une entreprise sont une fonction croissante de ses dépenses de publicité PUB, mais au fur et à mesure que les dépenses de publicité augmentent, l'accroissement des ventes devient de plus en plus faible, d'autant plus que le niveau de départ des dépenses publicitaires est élevé.

La relation entre les ventes *V* et les dépenses de publicité *PUB* est-elle bien représentée par une des spécifications suivantes, et laquelle ?

$$V_t = \beta_1 + \beta_2 PUB_t\,, \qquad \text{avec } \beta_1 > 0 \text{ et } \beta_2 > 0$$
$$V_t = \beta_1 PUB_t^{\beta_2}\,, \qquad \text{avec } \beta_1 > 0 \text{ et } 0 < \beta_2 < 1$$
$$V_t = \ln(\beta_1 + PUB_t^{\beta_2})\,, \qquad \text{avec } \beta_1 > 0 \text{ et } 0 > \beta_2 > -1$$

Solution

La deuxième spécification représente bien la relation entre ventes et dépenses publicitaires. La dérivée de V_t par rapport à PUB_t vaut en effet $\beta_1\beta_2 PUB_t^{\beta_2-1}$ et cette dérivée diminue quand PUB_t augmente, parce que $\beta_1 > 0$ et $0 < \beta_2 < 1$.

EXERCICE 2 ÉLASTICITÉ DES VENTES AUX PRIX

Énoncé

Pour que l'élasticité des ventes V au prix P du produit soit en valeur absolue une fonction décroissante des dépenses de publicité PUB, il faut qu'une des relations suivantes prévale :

$$V_t = \beta_1 P_t^{-(\beta_2+(\beta_3/PUB_t))}\,, \qquad \text{avec } \beta_1 > 0 \text{ et } \beta_2 > 0 \text{ et } \beta_3 > 0$$
$$V_t = \beta_1 + \beta_2 PUB_t + \beta_3 P_t\,, \qquad \text{avec } \beta_1 > 0 \text{ et } \beta_2 > 0 \text{ et } \beta_3 < 0$$
$$V_t = \beta_1 P_t^{\beta_2} PUB_t^{\beta_3}\,, \qquad \text{avec } \beta_1 > 0 \text{ et } \beta_2 < 0 \text{ et } \beta_3 > 0$$

Parmi ces trois relations, laquelle est à retenir ?

Solution

La première spécification est appropriée, puisque l'élasticité des ventes au prix vaut $-\big(\beta_2 + (\beta_3/PUB_t)\big)$. Cette expression est négative puisque $\beta_2 > 0$, $\beta_3 > 0$, et $PUB_t > 0$ par définition. En valeur absolue, cette élasticité vaut donc $\big(\beta_2 + (\beta_3/PUB_t)\big)$. Elle décroît si PUB_t augmente étant donné que PUB_t se trouve au dénominateur et que $\beta_3 > 0$.

Exercices

EXERCICE 3 SPÉCIFICATION D'UNE FONCTION DE PRODUCTION

Énoncé

Soit une fonction de production, qu'on représente de la manière suivante, et en prenant comme hypothèse d'absence de progrès technique :

$$Y = f(K, L)$$

Y est la quantité produite, K est le capital et L l'emploi.

Parmi ces deux spécifications, laquelle est réaliste : $Y = \alpha + \beta K + \gamma L$ ou $Y = AK^{\beta}L^{\gamma}$?

Solution

Exprimer cette relation sous une forme linéaire, $Y = \alpha + \beta K + \gamma L$, revient à imposer arbitrairement que les productivités marginales sont constantes, donc qu'elles ne dépendent pas des quantités de facteurs. En effet, avec une telle spécification, $\beta = \dfrac{\partial Y}{\partial K}$ est la productivité marginale du capital et $\gamma = \dfrac{\partial Y}{\partial L}$ est la productivité marginale du travail. Ces productivités marginales sont supposées indépendantes des quantités de facteurs, puisqu'elles sont égales à des constantes β et γ. Par conséquent, en modélisant la production de cette manière, on ignore délibérément des caractéristiques bien connues de beaucoup de processus de production réels. On ne tient pas compte en particulier des deux phénomènes suivants : la productivité marginale du travail diminue quand la quantité de travail augmente et que le stock de capital reste à un niveau constant, et elle croît quand le stock de capital augmente et que la quantité de travail reste inchangée. Concrètement, ajouter l'un après l'autre des ouvriers supplémentaires à une équipe qui travaille sur une machine conduit en général à des accroissements de moins en moins importants de la production et devient au bout d'un certain temps contre-productif (on provoque une congestion qui diminue la production). Par contre, mieux équiper les ouvriers permet d'augmenter la contribution productive apportée par une éventuelle main-d'oeuvre supplémentaire. Une spécification linéaire de la fonction de production n'implique pas ces propriétés réalistes ; elle est donc inadéquate dans le cas d'une fonction de production.

Pour représenter correctement de telles caractéristiques, on utilise des formes fonctionnelles non linéaires comme la fonction de Cobb-Douglas :

$$Y = AK^{\beta}L^{\gamma}$$

La productivité marginale du travail se mesure alors par $\dfrac{\partial Y}{\partial L} = A\gamma K^{\beta}L^{\gamma-1} = \dfrac{\gamma Y}{L}$ et la productivité marginale du capital par $\dfrac{\partial Y}{\partial K} = A\beta K^{\beta-1}L^{\gamma} = \dfrac{\beta Y}{K}$. Cette fois, les productivités marginales ne sont pas constantes, mais varient en fonction du niveau déjà atteint par L et K. On peut vérifier qu'elles respectent les propriétés réalistes mises en évidence précédemment. En effet, quand K est inchangé, le supplément $\dfrac{\partial Y}{\partial L}$ de production induit par l'intervention d'un ouvrier supplémentaire diminue au fur et à mesure qu'augmente le nombre d'ouvriers L déjà en place : $\dfrac{\partial\left(\dfrac{\partial Y}{\partial L}\right)}{\partial L} = A\gamma\,(\gamma - 1)\,K^{\beta}L^{\gamma-2} = \dfrac{\gamma\,(\gamma - 1)\,Y}{L^{2}}$ est

négatif à condition que γ soit inférieur à 1. Pour un nombre donné L d'ouvriers déjà à l'œuvre, augmenter le stock de capital K permet d'accroître le supplément de production apporté par un intervenant supplétif :

$$\frac{\partial\left(\frac{\partial Y}{\partial L}\right)}{\partial K} = A\gamma\beta K^{\beta-1}L^{\gamma-1} = \frac{\gamma\beta Y}{LK} > 0$$

EXERCICE 4 FONCTION DE CONSOMMATION À PRIX COURANTS OU CONSTANTS ?

Énoncé

Soient c_t les quantités consommées et p_t leur prix, à la période t. La consommation réelle est donc c_t et la consommation nominale $C_t = p_t c_t$. Le revenu nominal est Y_t et le revenu réel est $y_t = Y_t/p_t$. En termes nominaux, la relation entre consommation et revenu est $C_t = a + bY_t$ tandis qu'en termes réels, elle s'écrit $c_t = a + by_t$.

La relation entre consommation et revenu doit-elle être spécifiée en termes réels ou nominaux?

Solution

En termes nominaux, la relation est $C_t = a + bY_t$, et donc $p_t c_t = a + bY_t$, ce qui implique que $c_t = (a/p_t) + b(Y_t/p_t)$, c'est-à-dire $c_t = (a/p_t) + by_t$. Si les prix p_t et le revenu nominal Y_t doublent simultanément, le revenu réel y_t reste inchangé. Toutefois, le terme a/p_t change. La relation $C_t = a + bY_t$ implique donc que les quantités consommées c_t diminuent lorsque les prix et le revenu nominal doublent simultanément.

Logiquement, si les prix doublent et que le revenu nominal double aussi, cela ne change rien au pouvoir d'achat des consommateurs ; les quantités achetées devraient rester inchangées. La relation en termes nominaux ne reflète donc pas un comportement rationnel de la part des consommateurs. Il faut lui préférer la relation en termes réels : $c_t = a + by_t$.

EXERCICE 5 CONSOMMATION, REVENU DISPONIBLE ET SALAIRE

Énoncé

On ajoute à la relation $c = \alpha + \beta yd + \gamma r$ d'autres variables explicatives susceptibles de contribuer à déterminer l'évolution de la consommation, en l'occurrence le niveau moyen des salaires w – parce que « lorsque les salaires augmentent, le revenu disponible augmente et la consommation s'élève ». On formule une nouvelle relation linéaire de la forme :

$$c = \alpha + \beta yd + \gamma r + \lambda w \qquad (a'')$$

Cette suggestion est-elle raisonnable ?

Solution

Cette suggestion n'est pas fondée. Cette relation est redondante et n'est pas correctement spécifiée. Pourquoi? Le coefficient λ est la dérivée partielle de c par rapport au salaire w. Il mesure donc la réaction de la consommation c à une variation des salaires w, le revenu disponible yd et le taux d'intérêt restant inchangés. Or on a voulu justifier l'apport de w en indiquant que ses variations provoquent une variation du revenu disponible yd,

et par là même de c. Manifestement, λ ne mesure pas ce type d'effet, mais plutôt un impact direct hypothétique de w sur c, quand yd est inchangé (donc un effet de w sur c, qui ne passerait pas par yd). C'est tout à fait différent de l'effet indirect qu'on voulait (inutilement) mettre en évidence – w influence yd, qui lui-même influence c – et qui est déjà pris en compte à travers yd dans l'équation ; il n'est donc pas nécessaire d'ajouter w. Le coefficient β mesure l'impact sur c d'une variation de yd, quelles que soient les causes de ce changement, y compris une variation de w. Pour ajouter w à l'équation (a′), et donc utiliser (a″), il faut être sûr que, indépendamment de son effet indirect *via* son influence sur yd, w influence aussi directement c, pour une autre raison (c'est seulement cet autre effet qui sera mesuré par son coefficient). Dans l'exemple utilisé, il est difficile de justifier économiquement une telle hypothèse. Pour expliquer les variations de c, il est donc inutile d'ajouter la variable explicative w quand elle n'exerce qu'un effet indirect sur c. Mais quand l'intérêt de l'étude porte effectivement sur la mesure de l'effet indirect de w sur c, et non sur une explication des variations de c, comment mesurer cet effet indirect ? Il faut spécifier une nouvelle relation dans le modèle, qui explique yd en fonction de w et de ses autres déterminants qu'on représente ici par une variable x : par exemple, $yd = \rho + \varphi w + \mu x$. L'effet indirect de w sur c est alors $\dfrac{\partial c}{\partial w} = \dfrac{\partial c}{\partial yd}\dfrac{\partial y}{\partial w} = \beta\varphi$; il est obtenu à partir des coefficients de deux équations différentes.

Exercice 6 Taux d'intérêt nominal ou réel ?

Énoncé

Le revenu nominal est Y_t à la période t, et Y_{t+1} à la période $t+1$. Soient c_t les quantités consommées, et p_t leur prix, à la période t. Soient c_{t+1} les quantités consommées, et p_{t+1} leur prix, à la période $t+1$. Les consommations nominales des deux périodes sont $C_t = p_t c_t$ et $C_{t+1} = p_{t+1}c_{t+1}$. Le taux d'intérêt nominal est R_t. Le taux d'intérêt réel est $r_t = \big((1+R_t)/(1+I_t)\big) - 1$, où I_t est le taux d'inflation : $I_t = (p_{t+1} - p_t)/p_t$. Les consommateurs choisissent les quantités consommées c_t et c_{t+1} sous la contrainte budgétaire nominale intertemporelle $C_{t+1} = (Y_t - C_t)(1+R_t) + Y_{t+1}$. On veut spécifier un modèle expliquant les quantités consommées c_t en fonction du revenu réel y_t, du revenu réel y_{t+1} et du taux d'intérêt.

Celui-ci doit-il être le taux d'intérêt nominal ou réel ?

Solution

La contrainte budgétaire nominale est encore $p_{t+1}c_{t+1} = (Y_t - p_t c_t)(1+R_t) + Y_{t+1}$, ce qui implique la relation suivante entre les quantités consommées : $c_{t+1} = \big[\big((Y_t/p_t) - c_t\big)/\big((1+R_t)/(p_{t+1}/p_t)\big)\big] + (Y_{t+1}/p_{t+1})$. Cette contrainte peut se réécrire ainsi : $c_{t+1} = \big[\big((Y_t/p_t) - c_t\big)/\big((1+R_t)/(p_{t+1}/p_t)\big)\big] + (Y_{t+1}/p_{t+1})$. Elle devient donc $c_{t+1} = (y_t - c_t)/(1+r_t) + y_{t+1}$, où $y_t = Y_t/p_t$ est le revenu réel à la période t, $y_{t+1} = Y_{t+1}/p_{t+1}$ est le revenu réel à la période $t+1$, r_t est le taux d'intérêt réel, défini par $r_t = (1+R_t)/(1+I_t) - 1$ et I_t est le taux d'inflation défini par $I_t = (p_{t+1} - p_t)/p_t$. La contrainte budgétaire ainsi exprimée montre que les choix des quantités consommées aux périodes t et $t+1$ sont influencés par les revenus réels aux périodes t et $t+1$ et par le taux d'intérêt réel. C'est donc le taux d'intérêt réel, et non le taux d'intérêt nominal, qui doit intervenir dans une fonction explicative des quantités consommées.

Exercice 7 Choix des données

Énoncé

La fonction de consommation suivante explique les dépenses c en fonction du revenu disponible yd et du taux d'intérêt r :

$$c = \alpha + \beta yd + \gamma r$$

Vous voulez vérifier la pertinence de ce modèle, c'est-à-dire vous assurer qu'il peut rendre compte des données observées. Expliquez dans quels cas vous utilisez :

- des données en séries temporelles ;
- des données en coupe instantanée ;
- des données panel.

Solution

- **Séries temporelles.** Si la fonction de consommation semble une bonne description de la manière dont la consommation agrégée d'un pays évolue d'une période à l'autre, en fonction de l'évolution temporelle du revenu et du taux d'intérêt, cette fonction peut être vue comme une relation stable et valable à toute période t. Soient c_t la consommation réelle agrégée durant la période t, yd_t le revenu disponible réel durant la période t, et r_t le taux d'intérêt moyen durant la période t. L'équation devient :

$$c_t = \alpha + \beta yd_t + \gamma r_t\,, \quad \text{pour tout } t$$

  Les coefficients α, β, γ ne sont pas indicés par t, contrairement aux variables c_t, yd_t et r_t. On les suppose constants dans le temps. La relation économique théorique représentée par la fonction de consommation est vraisemblablement une loi économique stable dans le temps. C'est une hypothèse forte, mais dans la mesure où la théorie économique a une quelconque validité pour expliquer les phénomènes économiques, on peut supposer l'existence des relations stables. Pour les vérifier, il faut estimer leurs coefficients à partir des observations historiques de c, yd et r, appelées « séries temporelles ».

- **Données en coupe instantanée.** Si l'équation semble plutôt décrire la manière dont différents agents (ici les consommateurs) déterminent leur consommation particulière en fonction de leur revenu disponible personnel et du taux d'intérêt, durant une période donnée, elle peut être vue comme une relation commune aux différents agents. Soient c_i la consommation de l'agent i, yd_i le revenu disponible de l'agent i et r_i le taux d'intérêt auquel l'agent i peut prêter ou emprunter. Le modèle devient :

$$c_i = \alpha + \beta yd_i + \gamma r_i\,, \quad \text{pour tout } i$$

  Les taux de réaction à ces variables, c'est-à-dire les coefficients β et γ, sont supposés les mêmes pour tous les agents, durant une période d'étude donnée. Pour vérifier cette relation, il faut la confronter à des observations concrètes de la consommation, du revenu et du taux d'intérêt pour un ensemble d'agents différents, durant une période précise. Ces observations sont les données en coupe instantanée. Remarque : quand le taux d'intérêt est le même pour tous les consommateurs, et prend donc une valeur r, son effet est dilué dans un terme constant commun représenté par $\alpha + \gamma r$ et le taux de réaction γ n'est pas identifiable.

- **Données panel.** Si l'équation semble décrire la manière dont la consommation varie d'une période à l'autre et diffère d'un agent à l'autre en fonction de l'évolution dans le temps du revenu et du taux d'intérêt et de leurs différences d'un agent à l'autre, la fonction de consommation peut être vue comme une relation stable et commune à tous, décrivant le comportement de tous les agents durant toutes les périodes. Soient c_{it} la consommation en termes réels de l'agent i durant la période t, yd_{it} le revenu disponible réel de l'agent i durant la période t, et r_{it} le taux d'intérêt pour l'agent i durant la période t. L'équation devient :

$$c_{it} = \alpha + \beta yd_{it} + \gamma r_{it}\,, \quad \text{pour tout } i \text{ et pour tout } t$$

On peut remplacer r_{it} par r_t quand on suppose que tous les agents ont le même taux d'intérêt. Pour mesurer et vérifier une telle relation, il faut la confronter à des observations du revenu et du taux d'intérêt pour un ensemble d'agents différents, sur des périodes différentes de la consommation. De telles données sont appelées « données panel ».

EXERCICE 8 SPÉCIFICATION D'UNE FONCTION DE CONSOMMATION DYNAMIQUE

Énoncé

Les théories macroéconomiques à fondements microéconomiques impliquent généralement que la consommation réelle agrégée est, à l'équilibre (à « long terme »), proportionnelle au revenu disponible réel agrégé et que la constante de proportionnalité est une fonction du taux de croissance d'équilibre du revenu disponible réel, du taux d'inflation d'équilibre et du taux d'intérêt d'équilibre. Spécifiez un modèle dynamique explicatif de la consommation agrégée, en veillant à ce que ses solutions d'équilibre respectent ce qui vient d'être dit.

Solution

La relation d'équilibre théorique peut se formuler ainsi :

$$C_t = AY_t\,, \quad \text{avec } A = f(g_Y, g_P, R)$$

où

Y_t est le revenu disponible réel agrégé à la période t ;

g_Y est le taux de croissance d'équilibre du revenu disponible réel ;

g_P est le taux d'inflation d'équilibre (ou taux de croissance d'équilibre des prix) ;

R est la valeur d'équilibre du taux d'intérêt ;

C_t est la consommation réelle agrégée à la période t.

Cette relation théorique s'écrit :

$$\ln(C_t) = A' + \ln(Y_t)\,, \quad \text{où } A' = \ln(A)$$

La théorie implique donc que, à l'équilibre, l'élasticité de la consommation au revenu est unitaire :

$$\frac{\partial C}{\partial Y}\frac{Y}{C} = \frac{\partial \ln(C)}{\partial \ln(Y)} = 1$$

Il faut commencer par estimer, à partir des données, un **modèle dynamique général** (MDG) qui contient les valeurs présentes et passées de chaque variable :

$$\begin{aligned}\ln C_t = {} & \beta_1 + \beta_2 \ln Y_t + \beta_3 \ln Y_{t-1} + \beta_4 \ln P_t \\ & + \beta_5 \ln P_{t-1} + \beta_6 R_t + \beta_7 R_{t-1} + \beta_8 \ln C_{t-1} + u_t\end{aligned}$$

Ce modèle dynamique capte tous les délais d'ajustement, effets d'habitude et autres inerties, tous les déséquilibres de court terme qui font que la consommation n'est pas, à chaque période, en relation d'équilibre avec ses déterminants. On peut écrire ce modèle dynamique général d'une autre manière, sachant que l'égalité se maintient si l'on soustrait la même quantité à gauche et à droite :

$$\begin{aligned}\ln C_t - \ln C_{t-1} = {} & \beta_1 + \beta_2 \ln Y_t + \beta_3 \ln Y_{t-1} + \beta_4 \ln P_t + \beta_5 \ln P_{t-1} \\ & + \beta_6 R_t + \beta_7 R_{t-1} + (\beta_8 - 1) \ln C_{t-1} + u_t\end{aligned}$$

Le membre de droite est évidemment inchangé si on ajoute et soustrait en même temps les mêmes éléments :

$$\begin{aligned}\ln C_t - \ln C_{t-1} = {} & \beta_1 + \beta_2 \ln Y_t - \beta_2 \ln Y_{t-1} + \beta_2 \ln Y_{t-1} + \beta_3 \ln Y_{t-1} \\ & + \beta_4 \ln P_t - \beta_4 \ln P_{t-1} + \beta_4 \ln P_{t-1} + \beta_5 \ln P_{t-1} \\ & + \beta_6 R_t - \beta_6 R_{t-1} + \beta_6 R_{t-1} + \beta_7 R_{t-1} \\ & + (\beta_8 - 1) \ln C_{t-1} + u_t\end{aligned}$$

Cela peut encore s'écrire de la manière suivante :

$$\begin{aligned}\ln C_t - \ln C_{t-1} = {} & \beta_1 + \beta_2 (\ln Y_t - \ln Y_{t-1}) + (\beta_2 + \beta_3) \ln Y_{t-1} \\ & + \beta_4 (\ln P_t - \ln P_{t-1}) + (\beta_4 + \beta_5) \ln P_{t-1} \\ & + \beta_6 (R_t - R_{t-1}) + (\beta_6 + \beta_7) R_{t-1} + (\beta_8 - 1) \ln C_{t-1} + u_t\end{aligned}$$

On obtient donc le **modèle dynamique général reparamétré** (MDGR) :

$$\begin{aligned}\ln C_t - \ln C_{t-1} = {} & \beta'_1 + \beta'_2 (\ln Y_t - \ln Y_{t-1}) + \beta'_3 \ln Y_{t-1} + \beta'_4 (\ln P_t - \ln P_{t-1}) \\ & + \beta'_5 \ln P_{t-1} + \beta'_6 (R_t - R_{t-1}) + \beta'_7 R_{t-1} + \beta'_8 \ln C_{t-1} + u_t\end{aligned}$$

Les relations entre les paramètres des deux équations MDG et MDGR sont :

$$\begin{gathered}\beta'_1 = \beta_1\,, \quad \beta'_2 = \beta_2\,, \quad \beta'_3 = \beta_2 + \beta_3\,, \quad \beta'_4 = \beta_4\,, \\ \beta'_5 = \beta_4 + \beta_5\,, \quad \beta'_6 = \beta_6\,, \quad \beta'_7 = \beta_6 + \beta_7\,, \quad \beta'_8 = (\beta_8 - 1)\end{gathered}$$

Les équations MDG et MDGR ne sont pas deux modèles différents, mais deux écritures, deux représentations différentes du même modèle dynamique. L'une implique l'autre !

Il faut ensuite rechercher la **solution d'équilibre stationnaire** du modèle dynamique, qui est une propriété de ce modèle. Elle se présente sous la forme d'une relation entre les variables qu'il implique lorsqu'elles sont constantes dans le temps, lorsqu'elles sont en équilibre stationnaire. Dans cet exercice, la solution d'équilibre stationnaire du modèle dynamique est la relation entre les variables impliquées simultanément par l'équation MDG (ou MDGR) et les hypothèses de stationnarité suivantes :

$$\begin{aligned}C_t &= C_{t-1} = \tilde{C}\,, \quad \forall t \\ Y_t &= Y_{t-1} = \tilde{Y}\,, \quad \forall t \\ R_t &= R_{t-1} = \tilde{R}\,, \quad \forall t \\ P_t &= P_{t-1} = \tilde{P}\,, \quad \forall t\end{aligned}$$

Pour trouver cette solution, il suffit donc d'intégrer ces hypothèses de stationnarité dans l'équation MDG. On obtient :

$$\ln \tilde{C} = \beta_1 + \beta_2 \ln \tilde{Y} + \beta_3 \ln \tilde{Y} + \beta_4 \ln \tilde{P} + \beta_5 \ln \tilde{P} + \beta_6 \tilde{R} + \beta_7 \tilde{R} + \beta_8 \ln \tilde{C}$$

Cela implique :

$$\ln \tilde{C} = \frac{(\beta_1) + (\beta_2 + \beta_3) \ln \tilde{Y} + (\beta_4 + \beta_5) \ln \tilde{P} + (\beta_6 + \beta_7)\, \tilde{R}}{1 - \beta_8}$$

Ce résultat est la solution d'équilibre stationnaire du modèle dynamique [(1)]. Il s'agit de la relation (et non d'un nouveau modèle) qu'il implique dans le cas particulier d'une situation d'équilibre stationnaire. Pour tester sur les données la théorie, il faut estimer le modèle dynamique et vérifier que ses coefficients sont tels que sa solution d'équilibre stationnaire est compatible avec la relation d'équilibre de la théorie. Celle-ci est donc vérifiée si les hypothèses suivantes ne sont pas rejetées :

$$\frac{(\beta_4 + \beta_5)}{1 - \beta_8} = 0 \quad \text{et} \quad \frac{(\beta_2 + \beta_3)}{1 - \beta_8} = 1$$

On peut objecter que, sur des données macroéconomiques caractérisées par une croissance continue, le concept d'équilibre stationnaire est peu pertinent. On peut aussi rechercher la solution de croissance équilibrée du modèle dynamique qui est une autre de ses propriétés. Cette solution est la relation entre les variables que le modèle implique lorsque toutes celles qui représentent le flux, les stocks et les prix croissent à un taux constant dans le temps (elles sont en croissance équilibrée) et que les variables relatives aux taux sont constantes dans le temps (elles sont en équilibre stationnaire puisqu'il est insensé qu'un taux d'intérêt augmente à un taux de croissance constant indéfiniment). Dans cet exercice, la solution de croissance équilibrée du modèle dynamique est la relation entre les variables impliquées simultanément par l'équation MDG (ou MDGR) et les hypothèses de croissance équilibrée suivantes :

$$\begin{aligned} \ln(C_t) - \ln(C_{t-1}) &= g_C\,, \quad \forall t \\ \ln(Y_t) - \ln(Y_{t-1}) &= g_Y\,, \quad \forall t \\ R_t = R_{t-1} &= \tilde{R}\,, \quad \forall t \\ \ln(P_t) - \ln(P_{t-1}) &= g_P\,, \quad \forall t \end{aligned}$$

Pour trouver cette solution, il est préférable d'utiliser la représentation MDGR du modèle dynamique. Si l'on intègre ces hypothèses de croissance équilibrée dans l'équation MDG, on trouve :

$$g_C = \beta'_1 + \beta'_2 g_Y + \beta'_3 \ln Y_{t-1} + \beta'_4 g_P + \beta'_5 \ln P_{t-1} + \beta'_7 R_{t-1} + \beta'_8 \ln C_{t-1_t}\,, \quad \forall t$$

Cela implique :

$$\ln C_{t-1} = \frac{g_C - \beta'_1 - \beta'_2 g_Y - \beta'_3 \ln Y_{t-1} - \beta'_4 g_P - \beta'_5 \ln P_{t-1} - \beta'_7 R_{t-1}}{\beta'_8}, \quad \forall t$$

Et donc :

$$\ln C_t = \frac{g_C - \beta'_1 - \beta'_2 g_Y - \beta'_4 g_P - \beta'_3 \ln Y_t - \beta'_5 \ln P_t - \beta'_7 R_t}{\beta'_8}, \quad \forall t$$

1. On trouve exactement la même solution en substituant les hypothèses de stationnarité dans l'équation MDGR. C'est logique puisque MDG et MDGR sont deux représentations différentes du même modèle.

Ou, de manière équivalente :

$$\ln C_t = \frac{\left(\beta_1 + \beta_2 g_Y + \beta_4 g_P - g_C\right) + (\beta_2 + \beta_3) \ln Y_t + (\beta_4 + \beta_5) \ln P_t + (\beta_6 + \beta_7)\, \tilde{R}}{1 - \beta_8}, \ \forall t$$

Ce résultat est la solution de croissance équilibrée du modèle dynamique. Il s'agit de la relation qu'il implique dans le cas particulier d'une situation de croissance équilibrée. Pour tester sur les données la théorie, il faut estimer le modèle dynamique et vérifier que ses coefficients sont tels que sa solution de croissance équilibrée est compatible avec le postulat de relation d'équilibre théorique. Cette relation est donc vérifiée si les restrictions ou hypothèses suivantes ne sont pas rejetées :

$$\frac{(\beta_4 + \beta_5)}{1 - \beta_8} = 0 \quad \text{et} \quad \frac{(\beta_2 + \beta_3)}{1 - \beta_8} = 1$$

Pour s'assurer que le modèle dynamique général respecte la théorie, c'est-à-dire pour que ses solutions d'équilibre stationnaire et de croissance équilibrée soient compatibles avec la relation d'équilibre de la théorie économique, il suffit d'imposer ces contraintes à ses paramètres. Si l'on procède ainsi, le modèle dynamique général devient le mécanisme à correction d'erreur, en l'occurrence :

$$\begin{aligned}\ln C_t - \ln C_{t-1} = {} & \beta_1 + \beta_2 (\ln Y_t - \ln Y_{t-1}) + (1 - \beta_8)(\ln Y_{t-1} - \ln C_{t-1}) \\ & + \beta_4 (\ln P_t - \ln P_{t-1}) + \beta_6 (R_t - R_{t-1}) + (\beta_6 + \beta_7) R_{t-1} + u_t\end{aligned}$$

On peut calculer ses solutions d'équilibre stationnaire et de croissance équilibrée :

$$\ln C_t = \frac{(\beta_1) + (\beta_6 + \beta_7)\, \tilde{R}}{1 - \beta_8} + \ln Y_t$$

$$\ln C_t = \frac{\left(\beta_1 + (\beta_2 - 1)\, g_Y + \beta_4 g_P\right) + (\beta_6 + \beta_7)\, \tilde{R}}{1 - \beta_8} + \ln Y_t$$

Elles sont bien compatibles avec la théorie macroéconomique. Quand C est proportionnel à Y, $g_C = g_Y$.

EXERCICE 9 Spécification d'un modèle dynamique de taux de change

Énoncé

La théorie de la parité des pouvoirs d'achat implique qu'à l'équilibre, le taux de change entre deux devises s'ajuste de manière à égaliser le coût d'acquisition d'un panier de biens dans les deux pays concernés, lorsque les biens sont exprimés dans une même devise. À court terme, le taux de change fluctue également en fonction d'autres variables, tel le différentiel de taux d'intérêts nominaux entre les deux pays. Spécifiez un modèle explicatif de l'évolution du taux de change e_{ij} entre les devises de deux pays i et j, qui soit approprié à court terme tout en étant compatible avec la théorie de la parité des pouvoirs d'achat à long terme. Supposez que les périodes sont annuelles et que la dynamique peut se réduire à des retards d'une période.

Solution Soit e_{ij} le taux de change défini comme le prix d'une unité de devise i exprimé en unités de devise j. Les variables explicatives suggérées par l'énoncé sont :

p_i, l'indice des prix dans le pays i, en devises du pays i ;

p_j, l'indice de prix dans le pays j, en devise du pays j ;

R_i, le taux d'intérêt nominal à court terme pour la devise du pays i ;

R_j, le taux d'intérêt nominal à court terme pour la devise du pays j.

En données annuelles, on suppose que la dynamique peut se réduire à une période. On spécifie le modèle dynamique général :

$$\begin{aligned}\ln(e_{ijt}) = \beta_1 + \beta_2 R_{it} + \beta_3 R_{it-1} + \beta_4 \ln(p_{it}) + \beta_5 \ln(p_{jt}) + \beta_6 \ln(p_{jt-1}) + \beta_7 \ln(p_{it-1})\\ + \beta_8 R_{jt} + \beta_9 R_{jt-1} + \beta_{10} \ln(e_{ijt-1}) + u_t\end{aligned}$$

Cela peut s'écrire ainsi :

$$\begin{aligned}\big(\ln(e_{ijt}) - \ln(e_{ijt-1})\big) = \beta_1 + \beta_2(R_{it} - R_{it-1}) + (\beta_2 + \beta_3)R_{it-1} + \beta_4\big(\ln(p_{it}) - \ln(p_{it-1})\big)\\ + \beta_5\big(\ln(p_{jt}) - \ln(p_{jt-1})\big) + (\beta_5 + \beta_6)\ln(p_{jt-1})\\ + (\beta_4 + \beta_7)\ln(p_{it-1}) + \beta_8(R_{jt} - R_{jt-1})\\ + (\beta_8 + \beta_9)R_{jt-1} + (\beta_{10} - 1)\ln(e_{ijt-1}) + u_t\end{aligned}$$

La théorie de la parité des pouvoirs d'achat suggère qu'à l'équilibre :

$$p_{it} e_{ijt} = k p_{jt}$$

où k est un facteur de proportionnalité tenant compte de différences éventuelles dans la composition et le choix de l'année de base des indices de prix des deux pays. Cette relation d'équilibre s'écrit encore :

$$\ln(e_{ijt}) = A + \ln(p_{it}) - \ln(p_{it})$$

où $A = \ln(k)$. La solution d'équilibre stationnaire du modèle dynamique général est :

$$\begin{aligned}\ln(e_{ijt}) = \big(\beta_1 + (\beta_2 + \beta_3)R_{it} + (\beta_4 + \beta_7)\ln(p_{it})\\ + (\beta_5 + \beta_6)\ln(p_{jt}) + (\beta_8 + \beta_9)R_{jt}\big)/(1 - \beta_{10})\end{aligned}$$

puisque $R_{it} = R_{it-1}$, $R_{jt} = R_{jt-1}$, $e_{ijt} = e_{ijt-1}$, $p_{it} = p_{it-1}$ et $p_{jt} = p_{jt-1}$. La solution de croissance équilibrée s'écrit :

$$\begin{aligned}\ln(e_{ijt}) = \big((\beta_1 + \beta_4 g_i + \beta_5 g_j - g_e) + (\beta_2 + \beta_3)R_{it} + (\beta_4 + \beta_7)\ln(p_{it})\\ + (\beta_5 + \beta_6)\ln(p_{jt}) + (\beta_8 + \beta_9)R_{jt}\big)/(1 - \beta_{10})\end{aligned}$$

où $g_e = \ln(e_{ijt}) - \ln(e_{ijt-1})$, $g_i = \ln(p_{it}) - \ln(p_{it-1})$, $g_j = \ln(p_{jt}) - \ln(p_{jt-1})$, $R_{it} = R_{it-1}$ et $R_{jt} = R_{jt-1}$. Pour être compatibles avec la relation théorique $\ln(e_{ijt}) = A + \ln(p_{it}) - \ln(p_{it})$, les coefficients de ces solutions d'équilibre doivent être contraints de sorte que $(\beta_5 + \beta_6)/(1 - \beta_{10}) = (\beta_4 + \beta_7)/(1 - \beta_{10}) = 1$. À l'équilibre, la relation théorique $\ln(e_{ijt-1}) = A + \ln(p_{it}) - \ln(p_{it})$ implique aussi que $g_e = g_j - g_i$. En imposant ces restrictions au modèle dynamique général, on obtient le mécanisme à correction d'erreur :

$$\begin{aligned}\big(\ln(e_{ijt}) - \ln(e_{ijt-1})\big) = \beta_1 + \beta_2(R_{it} - R_{it-1}) + (\beta_2 + \beta_3)R_{it-1} + \beta_4\big(\ln(p_{it}) - \ln(p_{it-1})\big)\\ + \beta_5\big(\ln(p_{jt}) - \ln(p_{jt-1})\big)\\ + (\beta_5 + \beta_6)\big(\ln(p_{jt-1}) - \ln(p_{it-1}) - \ln(e_{ijt-1})\big)\\ + \beta_8(R_{jt} - R_{jt-1}) + (\beta_8 + \beta_9)R_{jt-1} + u_t\end{aligned}$$



Références bibliographiques

[JOH 1997] J. Johnson, J. DiNardo, *Econometric Methods*, McGraw Hill, 1997.

[HEN 1995] D. Hendry, *Dynamic Econometrics*, Oxford University Press, 1995.

[SPA 1986] A. Spanos, *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge, 1986.



Modèle linéaire en univers stationnaire

Modèle linéaire en univers stationnaire

Ce chapitre étudie les problèmes de spécification, d'estimation et d'inférence relatifs à des relations linéaires entre des processus stochastiques purement stationnaires ou stationnaires autour d'une tendance déterministe (voir chapitre 1). L'inférence statistique étudiée dans ce chapitre ne s'applique donc pas à des relations linéaires entre des processus stochastiques intégrés, c'est-à-dire non stationnaires à tendance stochastique ou à racine unitaire (voir chapitre 1). Les exercices proposés se concentrent sur l'estimation des relations linéaires par moindres carrés ordinaires au moyen des logiciels Excel, TSP, SPSS et Easyreg.

1 Présentation générale

1.1 Définition générale et notation générale

Un **modèle linéaire** à une équation, en séries temporelles, suppose qu'une variable aléatoire univariée Y_t est une fonction linéaire d'autres variables aléatoires univariées $X_{2t}, X_{3t} \ldots X_{kt}$, à laquelle s'ajoute une variable aléatoire univariée u_t appelée « terme d'erreur », et émet certaines hypothèses sur la distribution de toutes ces variables. On représente cet ensemble d'hypothèses de la manière suivante :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t\,, \quad \text{pour tout } t = 1 \ldots n \qquad (2.1)$$

avec $\beta_1, \beta_2 \ldots \beta_k$ des coefficients non aléatoires constants dans le temps.

- La variable Y_t est dite **dépendante**. Les variables $X_{2t}, X_{3t} \ldots X_{kt}$ sont dites **explicatives**. Y_t et les X_{it} sont des variables **aléatoires** (1) (au sens où chacune d'elles peut avoir, d'un point de vue conceptuel, plusieurs valeurs possibles, en fonction du hasard, même si l'on observe effectivement une seule réalisation puisque la période t n'« a lieu » qu'une fois) et **observables** (puisqu'on peut en observer une réalisation).

  La variable u_t est appelée **terme d'erreur** ou **perturbation**. C'est une variable **aléatoire** et **non observable** (en effet, $u_t = Y_t - \beta_1 - \beta_2 X_{2t} - \beta_3 X_{3t} - \cdots - \beta_k X_{kt}$, mais on ne peut pas déduire sa réalisation des réalisations observées de Y_t et des X_{it} car les coefficients β_i sont inconnus et donc non observés).

  $Y_t, X_{2t}, X_{3t} \ldots X_{kt}$ et u_t sont les composantes à la date t des processus stochastiques correspondants $Y = \{Y_\tau\}_{\tau=1\ldots n}, X_2 = \{X_{2\tau}\}_{t=1\ldots n}, X_3 = \{X_{3\tau}\}_{\tau=1\ldots n} \ldots X_k = \{X_{k\tau}\}_{t=1\ldots n}$, et $u = \{u_\tau\}_{\tau=1\ldots n}$.
- Le coefficient β_1 est souvent appelé **constante** (ou « terme constant ») du modèle linéaire. De façon implicite, une variable explicative X_{1t} vaut 1 à chaque période t ($X_{1t} = 1$ pour tout t), ce qui implique que $\beta_1 X_{1t} = \beta_1$ pour tout t. Pour cette raison, l'indice des variables explicatives commence à 2 dans la formulation du modèle. On compte donc k variables explicatives, constante comprise (soit $k-1$ sans la constante !). Les coefficients β_i sont des concepts **non aléatoires** (des valeurs uniques supposées exister dans la nature) et **non observables** (leur valeur est inconnue).

Exemple

La fonction de consommation suivante est un exemple de modèle linéaire :

$$\ln(C_t) = \beta_1 + \beta_2 \ln(YI_t) + \beta_3 R_t + u_t\,, \quad \text{pour tout } t = 1 \ldots n$$

$\ln(C_t)$ est le logarithme népérien de la consommation à prix constants C_t.
$\ln(YI_t)$ est le logarithme népérien du revenu disponible réel YI_t.
R_t représente le taux d'intérêt réel.
$\ln(C_t)$ correspond à la variable dépendante Y_t.
$\ln(YI_t)$ correspond à la variable explicative X_{2t}.
R_t correspond à la variable explicative X_{3t}.

On compte donc trois variables explicatives, constante comprise ($k = 3$). $\ln(C_t)$, $\ln(YI_t)$, R_t et u_t sont les composantes à la date t des processus stochastiques correspondants $\ln(C) = \{\ln(C_\tau)\}_{\tau=1\ldots n}$, $\ln(YI) = \{\ln(YI_\tau)\}_{\tau=1\ldots n}$, $R = \{R_\tau\}_{\tau=1\ldots n}$ et $u = \{u_t\}_{\tau=1\ldots n}$.

1. Pour quelques rappels utiles sur les concepts de variable aléatoire, on peut se référer utilement au livre de Patrick Roger, Probabilités, statistique et processus stochastiques, publié chez Pearson Education France dans la même collection.



Exemple

La fonction de consommation suivante est un autre exemple de modèle linéaire :

$$\ln(C_t) = \beta_1 + \beta_2 \ln(YI_t) + \beta_3 R_t + \beta_4 \ln(C_{t-1}) + \beta_5 \ln(YI_{t-1}) + u_t\,, \quad \text{pour tout } t = 2 \ldots n$$

$\ln(C_t)$ correspond à la variable dépendante Y_t.
$\ln(YI_t)$ correspond à la variable explicative X_{2t}.
R_t correspond à X_{3t}.
$\ln(C_{t-1})$ correspond à X_{4t}.
$\ln(YI_{t-1})$ correspond à X_{5t}.

On compte donc cinq variables explicatives, constante comprise ($k = 5$). Dans cet exemple, $X_{4t} = Y_{t-1}$ et $X_{5t} = X_{2t-1}$. $\ln(C_t)$, X_{2t}, R_t et u_t sont les composantes à la date t des processus stochastiques correspondants $\ln(C) = \{\ln(C_\tau)\}_{\tau=1\ldots n}$, $\ln(YI) = \{\ln(YI_\tau)\}_{\tau=1\ldots n}$, $R = \{R_\tau\}_{\tau=1\ldots n}$ et $u = \{u_\tau\}_{\tau=1\ldots n}$. $\ln(C_{t-1})$ est la composante de date $t-1$ du processus stochastique $\{\ln(C_\tau)\}_{\tau=1\ldots n}$. $\ln(YI_{t-1})$ est la composante de date $t-1$ du processus stochastique $\{\ln(YI_\tau)\}_{\tau=1\ldots n}$.

1.2 NOTATION MATRICIELLE

L'hypothèse d'un modèle linéaire reliant des processus stochastiques $X_2, X_3 \ldots X_k$ peut encore être présentée de la manière suivante :

$$Y = X\beta + u \tag{2.2}$$

où $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & \cdots & X_{kn} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$

Y est un vecteur à n éléments (une matrice $n \times 1$), X est une matrice à n lignes et k colonnes, β est un vecteur à k éléments (une matrice $k \times 1$) et u est un vecteur à n éléments (une matrice $n \times 1$). Exprimé en notation matricielle, le modèle linéaire implique que :

$$\begin{aligned} Y_1 &= \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \beta_4 X_{41} + \cdots + \beta_k X_{k1} + u_1 \\ Y_2 &= \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \beta_4 X_{42} + \cdots + \beta_k X_{k2} + u_2 \\ &\ldots \\ Y_n &= \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \beta_4 X_{4n} + \cdots + \beta_k X_{kn} + u_n \end{aligned}$$

L'équation (2.2) est bien équivalente à l'équation (2.1).

2 Interprétations du modèle linéaire et hypothèses sur les erreurs

Dans l'équation (2.1), la somme $\beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_k X_{kt}$ est souvent interprétée, de manière conventionnelle, comme la partie de la variable dépendante Y_t qui peut « s'expliquer » linéairement en fonction des variables explicatives $X_{2t}, X_{3t} \ldots X_{kt}$, tandis que le terme d'erreur u_t est interprété comme la partie ne pouvant « s'expliquer » linéairement en fonction des variables explicatives.

Pour intuitive qu'elle soit, cette interprétation est encore trop vague et manque de rigueur. Il reste à préciser les différents statuts statistiques d'un modèle linéaire. La section suivante établit le lien entre ces statuts et certaines propriétés du terme d'erreur.

2.1 Hypothèses sur le lien entre terme d'erreur et variables explicatives

Modèle linéaire comme approximation linéaire

Statistiquement, les coefficients d'un modèle linéaire sont les coefficients de l'approximation linéaire de la variable dépendante par les variables explicatives, à condition que l'espérance du terme d'erreur soit nulle et que les covariances entre le terme d'erreur et chaque variable explicative soient nulles. Cela se formalise ainsi : si les processus stochastiques $Y, X_2 \ldots X_k$ sont tels qu'ils sont reliés par le modèle linéaire (2.1) (voir section 1.1) sous les hypothèses que $E(u_t) = 0 \quad \forall t$ et $\text{Cov}(u_t, X_{it}) = 0 \quad \forall i = 2 \ldots k$ et $\forall t$, la partie $\beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_k X_{kt}$ représente l'**approximation linéaire** de la variable aléatoire Y_t par les variables aléatoires $X_{2t} \ldots X_{kt}$. En d'autres termes, $\beta = \Sigma_{X_t}^{-1} \Sigma_{X_t Y_t}$ où X_t est un vecteur [1] aléatoire à k éléments défini par $X_t = (1 X_{1t} \ldots X_{kt})'$.

Remarque

L'approximation linéaire d'une variable aléatoire Y_t par des variables aléatoires $X_{2t}, X_{3t} \ldots X_{kt}$ est une fonction linéaire des variables $X_{2t}, X_{3t} \ldots X_{kt}$, dont les coefficients sont choisis de manière à minimiser l'espérance du carré de l'écart entre Y_t et cette fonction linéaire.

Avec de telles hypothèses sur le terme d'erreur, le modèle linéaire suppose donc la constance dans le temps des coefficients de l'approximation linéaire de Y_t par $X_{2t}, X_{3t} \ldots X_{kt}$, et par là même un comportement particulier des espérances et variances des variables Y_t et X_{it} $(i = 2 \ldots k)$ ainsi que des covariances entre ces variables, aux différentes périodes t. En effet, si ces espérances, variances et covariances varient dans le temps, elles doivent le faire de manière telle que $\Sigma_{X_t}^{-1} \Sigma_{X_t Y_t}$ soit constante dans le temps.

Exemple avec une variable explicative ($k = 2$)

Le modèle linéaire $Y_t = \beta_1 + \beta_2 X_2 t + u_t$, lorsqu'il comprend les hypothèses $\text{Cov}(u_t, X_{2t}) = 0$ pour tout t et $E(u_t) = 0$ pour tout t, implique que la valeur de l'expression $E(Y_t) - \dfrac{\text{Cov}(Y_t, X_{2t})}{V(X_{2t})} E(X_{2t})$ et celle de l'expression $\dfrac{\text{Cov}(Y_t, X_t)}{V(X_{2t})}$ ne changent pas quelle que soit la période t. Les coefficients du modèle linéaire sont alors définis statistiquement ainsi :

$$\beta_1 = E(Y_t) - \frac{\text{Cov}(Y_t, X_{2t})}{V(X_{2t})} E(X_{2t}) \quad \text{et} \quad \beta_2 = \frac{\text{Cov}(Y_t, X_{2t})}{V(X_{2t})}$$

1. Σ_{X_t} est une matrice carrée à k lignes et k colonnes puisque c'est la matrice de variances et covariances du vecteur X_t. $\Sigma_{X_t Y_t}$ est une matrice à k ligne et 1 colonne, donc un vecteur, puisque c'est la matrice des covariances entre le vecteur X_t et la variable univariée Y_t.



Modèle linéaire comme approximation conditionnelle

La partie expliquée de la variable dépendante d'un modèle linéaire est aussi une approximation conditionnelle qui dépend d'une hypothèse plus forte que celle de la covariance nulle entre terme d'erreur et variables explicatives : il s'agit de l'indépendance entre terme d'erreur et variables explicatives. Si les processus stochastiques $Y, X_2 \dots X_k$ sont tels qu'ils sont reliés par le modèle (2.1) sous les hypothèses que $E(u_t) = 0 \ \forall t$ et que u_t est **indépendant** de $X_{it} \quad \forall i = 2 \dots k$ et $\forall t$, la somme $\beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_k X_{kt}$ représente l'**approximation conditionnelle** [1] de la variable aléatoire Y_t par les variables aléatoires $X_{2t} \dots X_{kt}$. Par conséquent :

$$\beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_k X_{kt} = E(Y_t | X_{2t} \dots X_{kt}) \qquad (2.3)$$

Il apparaît clairement que ce cas implique celui de la rubrique précédente (*Modèle linéaire comme approximation linéaire*). Un modèle linéaire, avec les hypothèses complémentaires que $E(u_t) = 0 \quad \forall t$ et que u_t est indépendant des variables explicatives $X_{2t} \dots X_{kt}$, peut donc être interprété de la manière suivante : l'approximation conditionnelle de Y_t par les variables explicatives $X_{2t} \dots X_{kt}$ est une fonction linéaire de ces variables explicatives avec des coefficients constants dans le temps. Cela revient à dire que l'approximation conditionnelle de Y_t par les variables explicatives est identique à l'approximation linéaire de Y_t par ces variables explicatives et que des coefficients sont constants dans le temps. D'ailleurs, l'hypothèse d'indépendance entre u_t et $X_{2t} \dots X_{kt}$ pour tout t implique que $\text{Cov}(u_t, X_{it}) = 0 \quad \forall i = 2 \dots k$ et $\forall t$ et donc que les β_i sont à la fois les coefficients de l'approximation linéaire et ceux de l'approximation conditionnelle.

Remarque

L'approximation conditionnelle est égale à l'approximation linéaire lorsque les variables dépendante et explicatives sont toutes distribuées normalement. En effet, si la fonction de densité jointe de Y_t et des X_{it} est une normale multivariée, l'espérance conditionnelle de Y_t, conditionnellement aux réalisations des X_{it}, est effectivement une fonction linéaire des réalisations des X_{it} et est égale à l'approximation linéaire. Toutefois, la normalité n'implique pas à elle seule que les coefficients de l'approximation conditionnelle (et linéaire) de Y_t en fonction des X_{it} soient constants dans le temps. Il faut, pour cela, que la distribution jointe ait des propriétés supplémentaires.

Autres cas

Dans beaucoup de cas, en raison du contexte (la problématique économique, financière ou marketing) dans lequel on suppose l'existence d'une relation linéaire constante entre des processus stochastiques $Y, X_2 \dots X_k$, du type de l'équation (2.1), on ne peut supposer que le terme d'erreur u est indépendant des variables explicatives, ou que les covariances entre le terme d'erreur et les variables explicatives sont toutes nulles. Dans ces conditions, l'expression $\beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_k X_{kt}$ n'est pas l'approximation conditionnelle ni l'approximation linéaire de Y_t par $X_{2t}, X_{3t} \dots X_{kt}$. Les coefficients β_i ne sont donc pas ceux définis pour les coefficients de l'approximation linéaire.

La **simultanéité** est une des causes principales de dépendance entre terme d'erreur et variables explicatives. Par simultanéité, on entend « influence réciproque entre variable dépendante et variables explicatives », c'est-à-dire « influences simultanées des variables

1. Pour quelques rappels utiles sur le concept d'espérance conditionnelle, le lecteur peut se référer au manuel de Patrick Roger, Probabilités, statistique et processus stochastiques, publié chez Pearson Education France dans la même collection.

explicatives sur la variable dépendante et de la variable dépendante sur certaines variables explicatives ».

Exemple

On suppose habituellement que la consommation agrégée C_t est fonction du revenu disponible YI_t :

$$C_t = \beta_1 + \beta_2 YI_t + u_t \quad \text{(a)}$$

Cette équation s'inscrit dans un contexte où le revenu disponible YI_t est défini comme une partie du PIB_t (produit intérieur brut), qui est la somme des valeurs ajoutées dégagées dans l'économie, soit approximativement la somme des revenus primaires versés sous la forme de salaires, dividendes, intérêts :

$$YI_t = PIB_t - IMP_t \quad \text{(b)}$$

où IMP est un montant d'impôts et de cotisations sociales nettes des prestations sociales octroyées.

$$PIB_t = C_t + I_t + G_t + X_t - M_t \quad \text{(c)}$$

I représente l'investissement avec les variations de stock, G les dépenses publiques, X les exportations et M les importations. Au regard de ces trois équations, on se rend compte que toute variation de u_t affecte C_t, qui, en variant, affecte PIB_t, qui, en variant, affecte YI_t. Le terme d'erreur u_t et la variable explicative YI_t ne sont donc pas indépendants :

$$u \underset{(a)}{\rightarrow} C \underset{(c)}{\rightarrow} PIB \underset{(b)}{\rightarrow} YI \quad \text{donc} \quad u \rightarrow YI$$

Il en résulte que les coefficients β_1 et β_2 ne sont pas les coefficients de l'approximation conditionnelle ni de l'approximation linéaire de C_t par YI_t. En particulier, β_2 n'est pas égal à $\mathrm{Cov}(C_t, YI_t)/V(YI_t)$.

2.2 HYPOTHÈSES POSSIBLES SUR L'ÉVOLUTION TEMPORELLE DU TERME D'ERREUR

Plusieurs hypothèses concernant la manière dont le terme d'erreur évolue dans le temps, et les liens éventuels qu'il a avec ses réalisations passées, sont envisageables et seule la nature de la problématique étudiée rend plausible l'une d'entre elles.

Terme d'erreur bruit blanc

Lorsqu'on fait l'hypothèse d'un modèle linéaire à coefficients constants reliant des processus stochastiques $Y, X_2 \dots X_k$, du type de l'équation (2.1), on se demande si l'ensemble des fluctuations systématiques de Y au cours du temps est « expliqué » par les fluctuations des variables explicatives $X_2 \dots X_k$. Si c'est le cas, toutes les composantes systématiques de Y_t sont prises en compte par la somme $\beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_k X_{kt}$ et le terme d'erreur u_t ne contient plus d'éléments systématiques. On peut donc supposer que u_t est un bruit blanc. En effet, les réalisations successives d'un bruit blanc forment une suite de valeurs de moyenne 0, d'amplitude ou de dispersion constante et sans lien linéaire entre elles. Il s'agit bien d'une succession de valeurs ne présentant aucun caractère systématique.

Remarque

Un bruit blanc est un processus stochastique dont la composante à chaque date a une espérance nulle et la même variance, et dont des composantes à des dates différentes ont une covariance nulle.

Terme d'erreur autocorrélé ou hétéroscédastique

Toutefois, dans beaucoup de cas, seule une partie des fluctuations systématiques de Y au cours du temps est « expliquée » par les fluctuations des variables explicatives $X_2 \ldots X_k$. L'autre partie se retrouve dans le terme d'erreur u. Celui-ci a donc une composante systématique, ce qui implique qu'il n'est pas un bruit blanc [1]. Trois cas de figure sont alors possibles :

- Soit le terme d'erreur d'une période est autocorrélé, c'est-à-dire qu'il est lié à toutes ou à certaines de ses valeurs passées : $\exists \theta \neq 0 | \operatorname{cov}(u_t, u_{t-\theta}) \neq 0$. On parle alors d'« autocorrélation du terme d'erreur » (ou des perturbations).
- Soit le terme d'erreur est hétéroscédastique, c'est-à-dire que sa variance (dispersion) varie dans le temps : $V(u_t) \neq \sigma_u^2 \quad \forall t$. On parle alors d'« hétéroscédasticité du terme d'erreur » (ou des perturbations).
- Soit il est à la fois autocorrélé et hétéroscédastique.

3 Estimation par la méthode des moindres carrés ordinaires

3.1 Principe de la méthode

On suppose que l'hypothèse d'un modèle linéaire à coefficients constants reliant des processus stochastiques $Y, X_2 \ldots X_k$ est correcte. On veut dire par là qu'il existe effectivement des coefficients « vrais inconnus » β_i constants dans le temps tels que :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_k X_{kt} + u_t, \quad \forall t, \quad \text{comme dans l'équation (2.1)}$$

ce que l'on peut encore représenter par :

$$Y = X\beta + u, \quad \text{comme dans l'équation (2.2)}$$

où Y, X, β et u sont les matrices définies à la section 1.3.

Plusieurs estimateurs du vecteur « vrai inconnu » β des coefficients « vrais inconnus » β_i du modèle linéaire sont possibles. Le choix du bon estimateur dépend du statut des coefficients β_i, c'est-à-dire de leur interprétation, qui dépend elle-même des propriétés des distributions de probabilité des variables aléatoires multivariées $Y, X_2 \ldots X_k$ et u (voir section 2). Dans cette section, on définit un estimateur très populaire : l'**estimateur des moindres carrés ordinaires**. On étudie ses propriétés sous différentes hypothèses sur le modèle linéaire sous-jacent, ce qui permet de déterminer dans quels cas le choix de cet estimateur est opportun et dans quels cas il ne l'est pas.

Il faut donc définir la formule d'un estimateur du vecteur « vrai inconnu » β, obtenue en appliquant le principe des moindres carrés ordinaires (MCO). Cet estimateur est noté $\hat{\beta}^{MCO}$. Il s'agit d'une formule à appliquer aux réalisations observées des processus

---

1. La problématique de la section 2.2 est indépendante de celle de la section 2.1 : si, par exemple, la partie expliquée de Y dans le modèle est l'approximation conditionnelle linéaire de Y et si, par là même, le terme d'erreur u est indépendant des X_i, cela n'implique pas pour autant que le terme d'erreur u soit un bruit blanc.

$Y, X_2 \dots X_k$, donc aux valeurs observées du vecteur Y et de la matrice X, pour obtenir une valeur estimée du vecteur β des coefficients « vrais inconnus ».

En définissant un estimateur particulier $\hat{\beta}^{MCO}$ de β, on définit forcément un **modèle estimé** du type :

$$Y = X\hat{\beta}^{MCO} + e \tag{2.4}$$

où $\hat{\beta}^{MCO}$ est l'estimateur de β et e est un résidu calculé, défini ainsi : $e = Y - X\hat{\beta}^{MCO}$.

Le modèle estimé (2.4) correspond au modèle linéaire $Y = X\beta + u$ de l'équation (2.2), où le vecteur des coefficients « vrais inconnus » β est remplacé par l'estimateur $\hat{\beta}^{MCO}$, qui est lui-même un vecteur, et où le vecteur des termes d'erreur « vrais inobservables » u est remplacé par le vecteur des résidus calculés e. On a en effet $\hat{\beta}^{MCO} = \begin{pmatrix} \hat{\beta}_1^{MCO} \\ \hat{\beta}_2^{MCO} \\ \vdots \\ \hat{\beta}_k^{MCO} \end{pmatrix}$ et $e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$. Par ailleurs, le vecteur $X\hat{\beta}^{MCO}$, noté $\hat{Y}$, est la partie expliquée de Y.

Le modèle estimé peut encore être présenté ainsi :

$$Y_t = \hat{\beta}_1^{MCO} + \hat{\beta}_2^{MCO} X_{2t} + \cdots + \hat{\beta}_k^{MCO} X_{kt} + e_t \,, \ \ \forall t = 1 \dots n \tag{2.5}$$

où, à la période t, $\hat{\beta}_1^{MCO} + \hat{\beta}_2^{MCO} X_{2t} + \cdots + \hat{\beta}_k^{MCO} X_{kt}$, noté $\hat{Y}_t$, est la partie expliquée de Y_t.

Le modèle estimé (2.5) correspond au modèle vrai (2.1) où les coefficients « vrais inconnus » β_i sont remplacés par leurs estimateurs $\hat{\beta}_i^{MCO}$ et où le terme d'erreur vrai u_t est remplacé par le résidu e_t. L'estimateur des moindres carrés ordinaires $\hat{\beta}^{MCO}$ minimise la somme des carrés des résidus, c'est-à-dire $e'e$. On résout donc le problème d'optimisation :

$$\min_{\hat{\beta}_1^{MCO}, \hat{\beta}_2^{MCO}, \dots, \hat{\beta}_k^{MCO}} \sum_{t=1}^{n} \left(Y_t - \hat{\beta}_1^{MCO} - \beta_2^{MCO} X_{2t} - \cdots - \beta_k^{MCO} X_{kt} \right)^2$$

Ou encore :

$$\min_{\hat{\beta}_1^{MCO}, \hat{\beta}_2^{MCO}, \dots, \hat{\beta}_k^{MCO}} \sum_{t=1}^{n} e_t^2$$

Ce problème d'optimisation peut être formulé de manière matricielle :

$$\min_{\hat{\beta}^{MCO}} \left(Y - X\hat{\beta}^{MCO} \right)' \left(Y - X\hat{\beta}^{MCO} \right) \quad \text{ou} \quad \min_{\hat{\beta}^{MCO}} e'e \tag{2.6}$$

La solution du problème d'optimisation est :

$$\hat{\beta}^{MCO} = (X'X)^{-1} X'Y \tag{2.7}$$

C'est la **formule de l'estimateur des moindres carrés ordinaires**. Chaque fois qu'un logiciel informatique (un simple tableur Excel ou un logiciel statistique ou économétrique sophistiqué) calcule par MCO une régression linéaire, il détermine les coefficients à l'aide de cette formule. L'ordinateur met les données de la variable dépendante dans un vecteur Y, et les données des variables explicatives dans une matrice X. Il calcule ensuite la formule (2.7).

Remarque

L'estimateur de MCO n'est possible que lorsque le rang de la matrice X est égal à k, c'est-à-dire quand aucune colonne de X (aucune variable explicative) n'est une combinaison linéaire exacte d'autres colonnes de X (d'autres variables explicatives). Si une telle combinaison existe, on parle alors de **multicolinéarité parfaite**. Il est alors impossible de distinguer quelle est la contribution spécifique de chaque variable à l'explication de la variable dépendante ; il est mathématiquement impossible de calculer la formule de l'estimateur. En effet $X'X$ est alors singulière et n'est pas inversible.

3.2 Propriétés algébriques de l'estimateur des moindres carrés ordinaires

Les propriétés algébriques de l'estimateur des moindres carrés ordinaires découlent de la méthode de calcul utilisée pour obtenir la formule de cet estimateur. Ce sont donc des propriétés *toujours vraies*, qui ne dépendent pas de la véracité de telle ou telle hypothèse sur la distribution de probabilité des variables impliquées dans le modèle linéaire « vrai » supposé. Ces propriétés sont les suivantes :

- La somme des résidus calculés vaut 0.

$$\sum_{t=1}^{n} e_t = 0 \tag{2.8}$$

- Les résidus calculés sont orthogonaux aux variables explicatives.

$$\sum_{i=1}^{n} X_{it} e_t = 0\,, \quad \text{pour tout } i = 1 \ldots k \tag{2.9}$$

et donc, de manière équivalente :

$$X_i' e = 0 \quad \text{pour tout } i = 1 \ldots k \tag{2.10}$$

où $X_i = (X_{1i} X_{2i} \ldots X_{ni})'$. Cette propriété s'exprime encore ainsi :

$$X' e = 0 \tag{2.11}$$

puisque les différentes colonnes de la matrice X sont les différentes variables X_i . Quand $i = 1$, cette propriété implique la précédente, vu que $X_{1t} = 1$ pour tout t. La propriété d'orthogonalité signifie que la direction vectorielle du vecteur e des résidus calculés est orthogonale à la direction vectorielle du vecteur X_i des observations de chaque variable explicative. Ce qui signifie que le vecteur Y a été partagé entre une partie dépendant des X_i et une partie résiduelle e qui n'a plus aucun lien algébrique linéaire avec les variables X_i .

- Les résidus calculés sont orthogonaux à la partie expliquée de la variable dépendante.

$$\sum_{i=1}^{n} \hat{Y}_t e_t = 0 \tag{2.12}$$

ou encore $\hat{Y}' e = 0$

• La somme des carrés des valeurs de la variable dépendante correspond à la somme des carrés des valeurs de la partie expliquée et à la somme des carrés des résidus calculés.

$$\sum_{t=1}^{n} Y_t^2 = \sum_{t=1}^{n} \hat{Y}_t^2 + \sum_{t=1}^{n} e_t^2 \tag{2.13}$$

ou encore $Y'Y = \hat{Y}'\hat{Y} + e'e$

Cette propriété peut encore s'exprimer sous forme d'écarts à la moyenne :

$$\sum_{t=1}^{n}(Y_t - \bar{Y})^2 = \sum_{t=1}^{n}(\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^{n} e_t^2 \tag{2.14}$$

où $\bar{Y} = \dfrac{\sum_{t=1}^{n} Y_t}{n} = \dfrac{\sum_{t=1}^{n} \hat{Y}_t}{n}$ est la moyenne algébrique des observations de la variable dépendante et est aussi égale à la moyenne algébrique des valeurs expliquées de la variable dépendante.

• On peut définir une mesure descriptive de la « qualité » de la régression linéaire calculée. Ce concept, appelé R^2, mesure la proportion (fraction, pourcentage...) de la variance empirique totale des observations de la variable dépendante qui est « expliquée », par la variabilité des variables explicatives.

$$R^2 = \frac{\sum_{t=1}^{n}(\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} = 1 - \frac{\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \tag{2.15}$$

On obtient le R^2 en divisant la variance empirique de la partie expliquée $\hat{Y}$ de la variable dépendante (cette variance empirique est $\dfrac{\sum_{t=1}^{n}(\hat{Y}_t - \bar{Y})^2}{n-1}$) par la variance empirique de la variable dépendante Y dans son entièreté (cette variance empirique est $\dfrac{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}{n-1}$). Bien entendu, $0 \leqslant R^2 \leqslant 1$. On montre que le R^2 est aussi égal au carré du coefficient de corrélation empirique entre le vecteur des observations de la variable dépendante et le vecteur des valeurs expliquées de la variable dépendante.

$$R^2 = \rho_{Y\hat{Y}}^2 = \left(\frac{\left(\dfrac{\sum_{t=1}^{n}(\hat{Y}_t - \bar{Y})\left(Y_t - \bar{Y}\right)}{n-1} \right)}{\sqrt{\dfrac{\sum_{t=1}^{n}(\hat{Y}_t - \bar{Y})^2}{n-1}}\sqrt{\dfrac{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}{n-1}}} \right)^2 \tag{2.16}$$

À partir de là, on peut interpréter le R^2, comme l'intensité de la liaison linéaire entre la variable dépendante et sa partie expliquée. Cette interprétation est bien sûr équivalente à la fraction de la variance empirique totale des observations de la variable dépendante expliquée par la variabilité des variables explicatives.

On montre que le R^2 augmente nécessairement lorsqu'on ajoute une variable explicative au modèle et qu'on l'estime de nouveau, car la somme des carrés des résidus diminue

nécessairement. C'est pour cela qu'a été inventé un autre concept, le R^2 ajusté, qui n'augmente que si l'introduction de la nouvelle variable induit une réduction « suffisamment » grande de la somme des carrés des résidus. Le R^2 ajusté se calcule par la formule suivante :

$$\bar{R}^2 = 1 - \frac{(n-1)\sum_{t=1}^{n} e_t^2}{(n-k)\sum_{t=1}^{n}(Y_t - \bar{Y})^2} = 1 - \frac{n-1}{n-k}\left(1 - R^2\right) = \frac{n-1}{n-k}R^2 - \frac{k-1}{n-k} \quad (2.17)$$

Le R^2 et le R^2 ajusté doivent être interprétés avec prudence. Une valeur élevée de ces statistiques n'implique pas nécessairement que le modèle est un « bon » modèle explicatif de la variable dépendante. Par exemple, si l'on régresse une série temporelle croissante sur une autre série temporelle croissante indépendante de la première, on obtient nécessairement un R^2 assez élevé, alors que la variable de droite n'influence pas, dans la réalité, les évolutions de la variable dépendante. Le R^2 ne fait que mesurer une corrélation, et celle-ci ne reflète pas nécessairement une véritable relation entre deux phénomènes.

3.3 Propriétés statistiques de l'estimateur des moindres carrés ordinaires

Cadre général

Les propriétés statistiques de l'estimateur des moindres carrés ordinaires portent sur la distribution de probabilité de l'estimateur considéré comme une variable aléatoire multivariée. Elles dépendent donc de la véracité de certaines hypothèses sur la distribution de probabilité des variables impliquées dans le modèle linéaire supposé. Les variables explicatives X_i et la variable dépendante Y sont aléatoires et multivariées. Comme l'estimateur $\hat{\beta}^{MCO}$ est égal à $(X'X)^{-1}X'Y$, il est une fonction de variables aléatoires et donc lui-même aléatoire ; les propriétés de la distribution de probabilité de $\hat{\beta}^{MCO}$ peuvent être déduites des propriétés de la distribution de probabilité conjointe de Y et des X_i. L'estimateur $\hat{\beta}^{MCO}$ peut encore être exprimé comme une fonction des X_i (donc de X, qui rassemble tous les X_i) et de u :

$$\hat{\beta}^{MCO} = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'u \quad (2.18)$$

Il en résulte que les propriétés de la distribution de probabilité de $\hat{\beta}^{MCO}$ peuvent aussi être déduites des propriétés de la distribution de probabilité conjointe de u et des X_i.

On peut s'intéresser d'une part aux propriétés de la distribution de probabilité conditionnelle de $\hat{\beta}^{MCO}$, conditionnellement aux réalisations observées des variables explicatives (ce que l'on note par $|X$), d'autre part aux propriétés de la distribution inconditionnelle (ou marginale) de $\hat{\beta}^{MCO}$.

Remarque

Les propriétés décrites dans les sections suivantes ne sont pas valables si la **variable dépendante retardée** (c'est-à-dire la variable dépendante prise à une date antérieure à la date t) est l'une des variables explicatives, pour deux raisons :

- Le concept de distribution conditionnelle à X n'aurait aucun sens [1].
- La distribution inconditionnelle (marginale) aurait un sens, mais n'aurait pas les mêmes propriétés qu'en l'absence de variable dépendante retardée parmi les variables explicatives.

Toutefois, les propriétés suivantes restent valables si des variables explicatives retardées sont présentes.

Terme d'erreur indépendant des variables explicatives

Si $E(u_t) = 0 \quad \forall t$ et si X_{it} est indépendant de $u_{t+\lambda} \quad \forall t = 1 \ldots n, \forall \lambda, \forall i = 2 \ldots k$, on a les propriétés suivantes :

a) Les estimateurs des moindres carrés ordinaires sont des estimateurs sans biais des coefficients « vrais inconnus » du modèle linéaire [2] :

$$E\left(\hat{\beta}^{MCO}|X\right) = \beta \text{ et, si } E\big((X'X)^{-1}X'\big) < \infty, \quad E\left(\hat{\beta}^{MCO}\right) = \beta \tag{2.19}$$

b) Les estimateurs des moindres carrés ordinaires sont des estimateurs convergents des coefficients « vrais inconnus » du modèle linéaire, à condition que $p\lim\left(\frac{X'L_nX}{n}\right) = Q$, où L_n est la matrice de variance et de covariance de u ($\Sigma_u = L_n$) [3] et où Q est une matrice finie et inversible :

$$p\lim \hat{\beta}^{MCO} = \beta \tag{2.20}$$

conditionnellement à la valeur réalisée de X et inconditionnellement. Ce résultat s'applique évidemment à chaque élément du vecteur estimateur (donc $p\lim \hat{\beta}_1^{MCO} = \beta_1, p\lim \hat{\beta}_2^{MCO} = \beta_2 \ldots p\lim \hat{\beta}_k^{MCO} = \beta_k$).

Remarques

Quand u est un bruit blanc, ce qui implique que $\Sigma_u = \sigma_u^2 I_n$, cette condition se réduit à l'hypothèse que $p\lim\left(\frac{X'X}{n}\right) = Q$ est finie et inversible. Cette hypothèse découle elle-même de l'hypothèse plus forte que les vecteurs X_t sont indépendants et identiquement distribués, avec $E(X_tX_t') = Q$ pour tout $t = 1 \ldots n$. Quant aux hypothèses que $E(u) = 0$ et $\Sigma_u = \sigma_u^2 I_n$, elles découlent elles-mêmes de l'hypothèse plus forte que les u_t sont indépendants et identiquement distribués avec une espérance 0 et une variance σ_u^2. La combinaison de ces hypothèses plus fortes sur X et u est souvent utilisée dans les manuels car elle implique que les variables X_tu_t

1. Si, par exemple, le modèle est $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 Y_{t-1} + u_t$ pour tout t, la variable Y_p (pour une période p quelconque) se trouve dans le membre gauche à la période p mais dans le membre droit à la période $p+1$. Elle se trouve donc à la fois dans Y et dans X. Il est alors impossible de raisonner conditionnellement à X puisque ce serait aussi raisonner conditionnellement à Y.

2. Donc $E\left(\hat{\beta}_1^{MCO}|X\right) = \beta_1, E\left(\hat{\beta}_2^{MCO}|X\right) = \beta_2 \ldots E\left(\hat{\beta}_k^{MCO}|X\right) = \beta_k$, et si $E\big((X'X)^{-1}X'\big) < \infty : E\left(\hat{\beta}_1^{MCO}\right) = \beta_1, E\left(\hat{\beta}_2^{MCO}\right) = \beta_2 \ldots E\left(\hat{\beta}_k^{MCO}\right) = \beta_k$.

3. Les éléments diagonaux de L_n sont les $V(u_t) = \sigma^2(t)$ pour $t = 1 \ldots n$, et les éléments « hors diagonale » sont les $\text{cov}(u_t, u_s) = v(t,s)$ pour $t, s = 1 \ldots n$ et $t \neq s$.



sont i.i.d. avec $E(X_t u_t) = 0$ et, compte tenu du théorème de Khinchine, que $p\lim\left(\frac{X'u}{n}\right) = 0$.

Quand on raisonne conditionnellement à la valeur réalisée de X, les hypothèses utilisées sont $\lim_{n\to\infty}\left(\frac{X'L_nX}{n}\right) = Q$ ou $\lim_{n\to\infty}\left(\frac{X'X}{n}\right) = Q$.

Interprétation du caractère sans biais de $\hat{\beta}^{MCO}$

Si on pouvait répéter plusieurs fois l'histoire économique de chaque période t, on aurait pour chaque période t plusieurs réalisations différentes $x_{2t}^1, x_{2t}^2 \ldots$ de la variable aléatoire univariée X_{2t}, plusieurs réalisations différentes, $x_{3t}^1, x_{3t}^2 \ldots$ de la variable aléatoire univariée $X_{3t} \ldots$ plusieurs réalisations différentes $x_{kt}^1, x_{kt}^2 \ldots$ de la variable aléatoire univariée X_{kt}, ainsi que plusieurs réalisations différentes $y_t^1, y_t^2 \ldots$ de la variable aléatoire univariée Y_t. On aurait ainsi plusieurs réalisations différentes $x_2^1, x_2^2, x_2^3 \ldots$ du processus stochastique X_2, plusieurs réalisations différentes $x_3^1, x_3^2, x_3^3 \ldots$ du processus stochastique $X_3 \ldots$ plusieurs réalisations différentes $x_k^1, x_k^2, x_k^3 \ldots$ du processus stochastique X_k, ainsi que plusieurs réalisations différentes $y^1, y^2, y^3 \ldots$ du processus stochastique Y. On aurait donc plusieurs réalisations $x^1, x^2, x^3 \ldots$ de la matrice X, et plusieurs réalisations différentes $y^1, y^2, y^3 \ldots$ du vecteur Y. On obtiendrait ainsi plusieurs réalisations $\hat{\beta}^{MCO^1}, \hat{\beta}^{MCO^2}, \hat{\beta}^{MCO^3} \ldots$ de $\hat{\beta}^{MCO}$, telles que $\hat{\beta}^{MCO^1} = (x^{1\prime}x^1)^{-1}x^{1\prime}y^1, \hat{\beta}^{MCO^2} = (x^{2\prime}x^2)^{-1}x^{2\prime}y^2, \hat{\beta}^{MCO^3} = (x^{3\prime}x^3)^{-1}x^{3\prime}y^3 \ldots$ Chacune de ces réalisations serait la valeur estimée obtenue lorsqu'on applique la formule de l'estimateur des moindres carrés ordinaires aux données d'un échantillon particulier. Le caractère sans biais de l'estimateur $\hat{\beta}^{MCO}$ implique que toutes ses réalisations différentes $\hat{\beta}^{MCO^1}, \hat{\beta}^{MCO^2}, \hat{\beta}^{MCO^3} \ldots$ seraient distribuées autour d'une valeur moyenne égale à β – le vecteur des valeurs « vraies inconnues » que l'on cherche à estimer – et n'implique donc pas que la valeur estimée obtenue en pratique pour un seul échantillon particulier soit égale à la valeur « vraie inconnue » du vecteur β, ni même qu'elle en soit proche. Le caractère sans biais de l'estimateur garantit seulement que la méthode d'estimation utilisée n'a pas une tendance systématique à sous-estimer ou à surestimer les coefficients « vrais inconnus ».

Qu'est-ce qui permet alors d'apprécier la précision de l'estimation d'un coefficient « vrai inconnu », obtenue avec la méthode des moindres carrés ordinaires sur un échantillon particulier ? En d'autres termes, comment savoir si la probabilité que la valeur estimée obtenue soit « proche » de la valeur « vraie inconnue » du coefficient est forte ou faible ? C'est la variance de l'estimateur des moindres carrés ordinaires.

Interprétation du caractère convergent de $\hat{\beta}^{MCO}$

Si la taille de l'échantillon d'observations (disponibles pour chaque série temporelle) devient très grande, la distribution de probabilité de l'estimateur $\hat{\beta}^{MCO}$ n'équivaut plus qu'à un point : on ne note plus de différences entre les valeurs estimées obtenues avec des échantillons différents. Ce point est précisément la valeur vraie du vecteur β que l'on cherche à estimer.

Terme d'erreur indépendant des variables explicatives et de type bruit blanc

Aux hypothèses de la section précédente (*Terme d'erreur indépendant des variables explicatives*), on ajoute l'absence d'hétéroscédasticité et d'autocorrélation du terme d'erreur.

Si u est bruit blanc – c'est-à-dire si $E(u_t) = 0 \quad \forall t$, que $V(u_t) = \sigma_u^2 \quad \forall t$, et que $\text{Cov}(u_t, u_{t-\theta}) = 0 \quad \forall t, \forall \theta \neq 0$ – et si X_{it} est indépendant de $u_{t+\lambda} \quad \forall t = 1 \ldots n, \forall \lambda$, $\forall i = 2 \ldots k$, on garde les propriétés a) et b) de la section précédente, et on a en plus les propriétés suivantes :

c) Calculée à partir des résidus de l'estimation du modèle par moindres carrés ordinaires, la formule $\hat{\sigma}_u^{2MCO} = \dfrac{e'e}{n-k}$ est un estimateur sans biais et convergent de la variance « vraie inconnue » du terme d'erreur :

$$E\left(\frac{e'e}{n-k}|X\right) = \sigma_u^2 \quad \text{et} \quad E\left(\frac{e'e}{n-k}\right) = \sigma_u^2 \tag{2.21a}$$

$$p\lim\left(\frac{e'e}{n-k}\right) = \sigma_u^2 \tag{2.21b}$$

d) Les estimateurs des moindres carrés ordinaires ont une matrice de variance et de covariance non observable, donnée par :

$$\Sigma_{\hat{\beta}^{MCO}|X} = \sigma_u^2\left(X'X\right)^{-1} \quad \text{et} \quad \Sigma_{\hat{\beta}^{MCO}} = \sigma_u^2 E\left(\left(X'X\right)^{-1}\right) \tag{2.22}$$

e) Un estimateur sans biais de cette matrice est fourni par la formule $\hat{\Sigma}_{\hat{\beta}^{MCO}} = \dfrac{e'e}{n-k}\left(X'X\right)^{-1}$:

$$E\left(\hat{\Sigma}_{\hat{\beta}^{MCO}}|X\right) = E\left(\frac{e'e}{n-k}\left(X'X\right)^{-1}|X\right) = \Sigma_{\hat{\beta}^{MCO}|X} \tag{2.23a}$$

$$E\left(\hat{\Sigma}_{\hat{\beta}^{MCO}}\right) = E\left(\frac{e'e}{n-k}\left(X'X\right)^{-1}\right) = \Sigma_{\hat{\beta}^{MCO}} \tag{2.23b}$$

f) Parmi tous les estimateurs sans biais de β qui sont des fonctions linéaires de la variable dépendante, ceux qui utilisent la méthode des moindres carrés ordinaires sont les plus précis (c'est le célèbre théorème de Gauss-Markov) :

Pour tout estimateur B tel que $E(B|X) = E(B) = \beta$ et tel qu'il existe une matrice A vérifiant $B = AY$, $\Sigma_{B|X} - \Sigma_{\hat{\beta}^{MCO}|X}$ est une matrice semi définie positive[1], de même que $\Sigma_B - \Sigma_{\hat{\beta}^{MCO}}$.

g) Les estimateurs de moindres carrés ont une distribution asymptotique normale, inconditionnellement ou conditionnellement aux valeurs réalisées des variables explicatives.

$$\sqrt{n}\left(\hat{\beta}^{MCO} - \beta\right) \xrightarrow{L} N\left(0, \sigma_u^2 Q^{-1}\right)$$

$$\text{si } p\lim\left(\frac{X'X}{n}\right) = Q \text{ est finie et de rang plein}^{(2)} \tag{2.24}$$

$$\sqrt{n}\left(\hat{\sigma}_u^{2MCO} - \sigma_u^2\right) \xrightarrow{L} N\left(0, \left(\frac{\mu_4}{\sigma_u^4} - 1\right)\sigma_u^2\right) \tag{2.25}$$

Les distributions marginales et conditionnelles ont les unes et les autres une distribution asymptotique normale.

---

1. Cela signifie en quelque sorte que la matrice $\Sigma_{B|X}$ est « plus grande » que la matrice $\Sigma_{\hat{\beta}^{MCO}|X}$. En effet, si un scalaire a est plus grand qu'un scalaire b (on note $a > b$), il existe nécessairement un scalaire c supérieur à 0 ($c > 0$) et tel que $a = b + c$ (par exemple $8 > 3$ car $8 = 3 + 5$ où $5 > 0$). On généralise ce raisonnement à des matrices : une matrice A est « plus grande » qu'une matrice B s'il existe une matrice C définie positive telle que $A = B + C$. Une matrice C est définie positive si pour tout vecteur x différent de 0, $x'Cx > 0$. Le théorème de Gauss-Markov établit l'efficience relative de l'estimateur de MCO. On ne peut établir l'efficience absolue qu'en posant des hypothèses supplémentaires sur la distribution des variables ou du terme d'erreur.

2. Quand on raisonne conditionnellement aux valeurs réalisées des variables explicatives, donc à la valeur réalisée de X, cette condition devient « si $\lim\limits_{n\to\infty}\left(\dfrac{X'X}{n}\right) = Q$ est finie et de rang plein ».

h) Une fonction de l'écart entre l'estimateur $\hat{\beta}_i^{MCO}$ d'un coefficient et la valeur « vraie inconnue » β_i de ce coefficient est distribuée asymptotiquement d'après une normale centrée réduite, inconditionnellement ou conditionnellement aux valeurs réalisées des variables explicatives :

$$\forall i = 1 \ldots k \quad \frac{\hat{\beta}_i^{MCO} - \beta_i}{S_{\hat{\beta}_i^{MCO}}} \xrightarrow{L} N(0,1) \tag{2.26}$$

où $S_{\hat{\beta}_i^{MCO}}$ est l'écart type estimé de $\hat{\beta}_i^{MCO}$, et donc la racine carrée du i^e élément de la diagonale principale de $\hat{\Sigma}_{\hat{\beta}^{MCO}} = \dfrac{e'e}{n-k}\left(X'X\right)^{-1}$.

Terme d'erreur indépendant des variables explicatives, de type bruit blanc et distribué normalement

Aux hypothèses de la section précédente (*Terme d'erreur indépendant des variables explicatives et de type bruit blanc*), on ajoute l'hypothèse de normalité du terme d'erreur.

Si u est un bruit blanc – c'est-à-dire si $E(u_t) = 0 \quad \forall t$, que $V(u_t) = \sigma_u^2 \quad \forall t$, que $\text{Cov}(u_t, u_{t-\theta}) = 0 \quad \forall t, \forall \theta \neq 0$ – si X_{it} est indépendant de $u_{t+\lambda} \quad \forall t = 1 \ldots n, \forall \lambda$, $\forall i = 2 \ldots k$, et si u est normal : $u \sim N(0, \sigma_u^2 I_n)$, on garde les propriétés a), b), c), d), e), f), g), h) de la section précédente, et on a en plus les propriétés suivantes :

i) Les estimateurs des moindres carrés ordinaires sont distribués normalement pour tout n, conditionnellement à la valeur réalisée de X :

$$\hat{\beta}^{MCO} \sim N\left(\beta, \sigma_u^2 \left(X'X\right)^{-1}\right) \quad \text{conditionnellement à } X \tag{2.27a}$$

Il s'agit d'une distribution exacte, quelle que soit la taille de l'échantillon, y compris si elle est petite.

Toutefois, ce résultat n'est valable que pour la distribution conditionnelle $f_{\hat{\beta}^{MCO}|X}\left(\hat{\beta}^{MCO}|X\right)$. La distribution inconditionnelle, $f_{\hat{\beta}^{MCO}}\left(\hat{\beta}^{MCO}\right)$, ne suit pas une distribution exacte $N\left(\beta, \sigma_u^2 E\left(\left(X'X\right)^{-1}\right)\right)$ sur de petits échantillons, même si le terme d'erreur a une distribution normale, mais elle est asymptotiquement normale (c'est un résultat qui ne nécessite pas une distribution normale du terme d'erreur). Conditionnellement à X, on dispose également d'une distribution exacte sur de petits échantillons pour l'estimateur de la variance du terme d'erreur :

$$\frac{(n-k)\hat{\sigma}_u^{2\,MCO}}{\sigma_u^2} = \frac{e'e}{\sigma_u^2} \sim \chi^2_{n-k} \tag{2.27b}$$

j) Une fonction de l'écart entre l'estimateur $\hat{\beta}_i^{MCO}$ d'un coefficient et la valeur « vraie inconnue » β_i de ce coefficient est distribuée d'après une loi de Student pour tout n, conditionnellement à la valeur réalisée de X ou inconditionnellement :

$$\forall i = 1 \ldots k \; : \; \frac{\hat{\beta}_i^{MCO} - \beta_i}{S_{\hat{\beta}_i^{MCO}}} \sim t_{n-k} \tag{2.28}$$

où $S_{\hat{\beta}_i^{MCO}}$ est l'écart type estimé de $\hat{\beta}_i^{MCO}$, et donc la racine carrée du i^e élément de la diagonale principale de $\hat{\Sigma}_{\hat{\beta}^{MCO}}$. Ce résultat est valable quelle que soit la valeur de n. Il s'agit bien d'une distribution exacte, même sur de petits échantillons.

k) Une fonction, F-statistic $= \left(\frac{R^2}{1-R^2}\right)\left(\frac{n-k}{k-1}\right)$, est distribuée[1] selon une $F(k-1, n-k)$ pour tout n, sous l'hypothèse nulle que tous les coefficients, sauf la constante β_1, sont nuls, conditionnellement à la valeur réalisée de X ou inconditionnellement. Ce résultat est valable quelle que soit la valeur de n. Il s'agit bien d'une distribution exacte, même sur de petits échantillons.

l) Une fonction $\frac{\left(\hat{\beta}^{MCO}-\beta\right)' X'X\left(\hat{\beta}^{MCO}-\beta\right)(n-k)}{(e'e)\,k}$) est distribuée selon une $F(k, n-k)$ pour tout n, conditionnellement à la valeur réalisée de X ou inconditionnellement. Ce résultat est valable quelle que soit la valeur de n. Il s'agit bien d'une distribution exacte, même sur de petits échantillons.

4 Modèle linéaire dynamique

4.1 Présentation générale

Un modèle linéaire est dit **dynamique** lorsque les k variables explicatives comprennent un ou plusieurs retards de la variable dépendante (le retard d'une variable est la valeur de cette variable à une date antérieure à t). Avec m retards de la variable dépendante, le modèle linéaire dynamique s'écrit de la manière suivante :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_{k-m} X_{(k-m)t} + \beta_{k-m+1} Y_{t-1} + \beta_{k-m+2} Y_{t-2} + \cdots + \beta_k Y_{t-m} + u_t \quad \forall t = m+1 \ldots n \qquad (2.29)$$

Cela peut s'écrire : $Y = X\beta + u$, où $Y = \begin{pmatrix} Y_{m+1} \\ Y_{m+2} \\ \vdots \\ Y_n \end{pmatrix}$,

$$X = \begin{pmatrix} 1 & X_{2(m+1)} & \cdots & X_{(k-m)(m+1)} & Y_m & Y_{m-1} & \cdots & Y_1 \\ 1 & X_{2(m+2)} & \cdots & X_{(k-m)(m+2)} & Y_{m+1} & Y_m & \cdots & Y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & \cdots & X_{(k-m)n} & Y_{n-1} & Y_{n-2} & \cdots & Y_{n-m} \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \quad u = \begin{pmatrix} u_{m+1} \\ u_{m+2} \\ \vdots \\ u_n \end{pmatrix}$$

Y est un vecteur à $n-m$ éléments, X une matrice à $n-m$ lignes et k colonnes, β un vecteur à k éléments et u un vecteur à $n-m$ lignes. L'estimateur des moindres carrés ordinaires de β est de nouveau $\hat{\beta}^{MCO} = (X'X)^{-1}X'Y$.

1. Une distribution $F(k-1, n-k)$ est une loi de Fischer à $k-1$ degrés de liberté au numérateur et à $n-k$ degrés de liberté au dénominateur.



4.2 Propriétés statistiques de l'estimateur de MCO

Le terme d'erreur u_t d'un modèle linéaire dynamique, à une période spécifique, n'est jamais indépendant de toutes les variables explicatives mises en jeu à toutes les périodes. Si, par exemple, le modèle exprime, à la période t, la variable dépendante Y_t en fonction de la variable explicative Y_{t-1}, cette dernière est nécessairement corrélée avec le terme d'erreur u_{t-1} du modèle à la période $t-1$.

Il en résulte que l'estimateur de MCO est nécessairement biaisé. Si le terme d'erreur n'est pas autocorrélé, donc s'il s'agit d'un bruit blanc, l'estimateur de MCO est toutefois convergent ; sinon, il ne l'est pas.

5 Tests de mauvaise spécification du modèle

5.1 Problématique des tests

On distingue principalement deux types de tests d'hypothèses dans le contexte du modèle linéaire :

- Ceux qui portent sur les valeurs des coefficients β_i ou σ_u du modèle estimé (voir chapitre 3). Ils vérifient des restrictions sur les paramètres du modèle en question.
- Ceux qui portent, entre autres, sur l'absence d'autocorrélation ou d'hétéroscédasticité du terme d'erreur u, ou sur la constance des coefficients dans le temps, etc. Ils sont appelés « tests de mauvaise spécification du modèle » et vérifient généralement des restrictions par rapport à un modèle plus général que celui estimé.

Exemple

Après avoir estimé un modèle $Y_t = \beta_1 + \beta_2 X_{2t} + u_t$, on souhaite tester l'hypothèse d'absence d'autocorrélation de u_t en vérifiant que $\rho = 0$ dans le modèle plus général $Y_t = \beta_1 + \beta_2 X_{2t} + u_t$, avec $u_t = \rho u_{t-1} + v_t$, où v_t est un bruit blanc.

Qu'il soit du premier type ou du deuxième, tout test est soit exact (on parle alors de « test de petit échantillon »), soit asymptotique. Quand il est exact, on connaît la fonction de densité, sous l'hypothèse nulle, quelle que soit la grandeur de l'échantillon utilisé, même s'il est petit. Quand il est asymptotique, on ne connaît la fonction de densité, sous l'hypothèse nulle, que pour une taille d'échantillon tendant vers l'infini. La fonction de densité utilisée pour un test asymptotique n'est donc une bonne approximation de la fonction « vraie inconnue » que lorsque l'échantillon est grand.

Tests exacts

On ne peut obtenir un test exact qu'en faisant une hypothèse de distribution exacte sur le terme d'erreur du modèle. Par exemple, si l'on pose que le terme d'erreur est normalisé et de type bruit blanc, $\dfrac{\hat{\beta}_i^{MCO} - \beta_i}{S_{\hat{\beta}_i^{MCO}}} \sim t_{n-k}$ quelle que soit la taille n de l'échantillon. Ce résultat fournit un test exact de l'hypothèse que β_i vaut une valeur particulière. Pour tester simultanément un ensemble d'hypothèses linéaires sur plusieurs coefficients du modèle, on dispose également de tests qui ont une distribution exacte F lorsque le terme d'erreur est normal (voir chapitre 3).

Toutefois, pour certaines hypothèses, les tests n'ont pas de distribution connue sur un petit échantillon, même si l'on admet la normalité des erreurs. On utilise alors les tests asymptotiques.

Tests asymptotiques

Test du rapport de vraisemblance

Pour tester une hypothèse au moyen d'un test du rapport de vraisemblance (dit « *LR* » pour « likelihood ratio »), on est obligé de faire des hypothèses sur la distribution exacte des variables aléatoires du problème, de manière à déduire la fonction de vraisemblance. Concernant le modèle linéaire, il faut généralement prendre l'hypothèse que u est distribué normalement.

La forme générale d'un test *LR* est $\lambda = \frac{L_0}{L_1}$, où L_0 est la valeur de la fonction de vraisemblance maximisée sous l'hypothèse nulle H_0, et L_1 est la valeur de la fonction de vraisemblance maximisée sous l'hypothèse H_1.

Parfois, il est possible de trouver la distribution exacte de λ sous l'hypothèse H_0, même pour de petits échantillons. C'est en général le cas quand λ est une fonction monotone croissante d'un test dont on connaît la fonction de densité exacte sous l'hypothèse H_0. On a alors des valeurs critiques exactes pour tester l'hypothèse H_0 au moyen du test λ.

Toutefois, le plus souvent, il faut se contenter de la distribution asymptotique de λ. On montre en effet que, si H_0 est vraie et si H_0 est emboîtée dans H_1 (c'est-à-dire que, sous l'hypothèse H_0, si le modèle est un cas particulier du modèle qui prévaut sous l'hypothèse H_1), $-2\ln(\lambda) \xrightarrow{L} \chi^2_r$, où r est le nombre de restrictions que H_0 représente sur les paramètres du modèle, par rapport à leur statut sous H_1. On obtient ce résultat sous les conditions de régularité qui sont nécessaires pour garantir la normalité asymptotique et l'efficience des estimateurs du maximum de vraisemblance. La distribution n'est pas la distribution exacte de $-2\ln(\lambda)$ sur un petit échantillon. Si on choisit une valeur critique $\chi^2_{r1-\alpha}$, α est le seuil de signification nominal du test, mais pas le vrai seuil, qui est lui inconnu (il correspond à la probabilité de rejeter H_0 alors qu'elle est vraie, la puissance du test P valant par définition un moins la probabilité d'accepter H_0 alors qu'elle est fausse, donc un moins le risque de deuxième espèce).

Si φ est le vecteur des coefficients du modèle, si $\hat{\varphi}$ est l'estimateur du maximum de vraisemblance sans contrainte, et si $\hat{\varphi}_0$ est l'estimateur du maximum de vraisemblance sous la contrainte H_0, on montre que $-2\ln\lambda \simeq \left(\hat{\varphi} - \hat{\varphi}_0\right)' \left.\frac{\partial^2 L}{\partial\varphi\partial\varphi'}\right|_{\varphi=\hat{\varphi}} \left(\hat{\varphi} - \hat{\varphi}_0\right)$ où $L = L(\varphi)$ est la fonction de vraisemblance.

La réalisation d'un test du rapport de vraisemblance exige d'estimer le modèle contraint sous H_0, et le modèle non contraint sous H_1. L'idée générale du test est la suivante : si l'hypothèse H_0 est vraie, on peut l'imposer à l'estimation avec une faible perte de vraisemblance : L_0 doit être proche de L_1, et donc λ doit être à peine plus petit que 1. Par conséquent, on rejette H_0 quand λ est bien plus petit que 1, et donc quand $-2\ln(\lambda)$ est trop grand.

Test de Wald

Pour tester une hypothèse au moyen d'un test de Wald (ou test W), il suffit d'estimer le modèle non contraint, c'est-à-dire sous l'hypothèse H_1. Le principe du test est le suivant : si

H_0 est vraie, les valeurs estimées des coefficients, obtenues sans imposer H_0 (donc sous H_1), doivent approximativement respecter les contraintes qui seraient imposées par H_0. On rejette donc H_0 si les valeurs estimées obtenues librement s'éloignent fortement des contraintes représentées par H_0.

On peut représenter un ensemble de m restrictions linéaires sur les coefficients d'un modèle statistique quelconque par l'équation $R\varphi = r$, où φ est le vecteur des p coefficients du modèle. La matrice R a m lignes et p colonnes tandis que le vecteur r a m éléments. De manière générale, si un estimateur convergent $\hat{\varphi}$ du vecteur de coefficients φ est distribué asymptotiquement de la manière suivante : $\sqrt{n}\left(\hat{\varphi} - \varphi\right) \xrightarrow{L} N\left(0, \Sigma\left(\varphi\right)^{-1}\right)$, alors $n\left(R\hat{\varphi} - r\right)'\left(R\Sigma\left(\varphi\right)^{-1}R'\right)^{-1}\left(R\hat{\varphi} - r\right) \xrightarrow{L} \chi^2_m$ sous l'hypothèse $R\varphi = r$. Si l'on dispose d'une matrice observable $\hat{\Sigma}\left(\hat{\varphi}\right)$ telle que $p\lim \dfrac{\hat{\Sigma}\left(\hat{\varphi}\right)}{n} = \Sigma\left(\varphi\right)$ et si les hypothèses $R\varphi = r$ sont vraies, $W = \left(R\hat{\varphi} - r\right)'\left(R\hat{\Sigma}\left(\hat{\varphi}\right)^{-1}R'\right)^{-1}\left(R\hat{\varphi} - r\right) \xrightarrow{L} \chi^2_m$. La statistique W est le test de Wald des hypothèses $R\varphi = r$.

Test du multiplicateur de Lagrange

Pour tester une hypothèse au moyen d'un test du multiplicateur de Lagrange (test *LM*, Lagrange Multiplier), il suffit d'estimer le modèle contraint, c'est-à-dire sous l'hypothèse H_0. Le principe du test est le suivant : si H_0 est vraie et qu'on maximise la vraisemblance en imposant H_0, les contraintes représentées par cette hypothèse induisent une faible perte de vraisemblance : les multiplicateurs de Lagrange associés aux contraintes imposées par H_0 doivent être proches de 0. On montre que si les hypothèses $R\varphi = r$ sont vraies, $LM = \left(\left.\dfrac{\partial^2 L}{\partial\varphi\partial\varphi'}\right|_{\varphi=\hat{\varphi}_0}\right)' I\left(\hat{\varphi}_0\right)^{-1}\left.\dfrac{\partial^2 L}{\partial\varphi\partial\varphi'}\right|_{\varphi=\hat{\varphi}_0} \xrightarrow{L} \chi^2_m$ où $I\left(\hat{\varphi}_0\right) = -E\left(\left.\dfrac{\partial^2 L}{\partial\varphi\partial\varphi'}\right|_{\varphi=\hat{\varphi}_0}\right)$. Ce test *LM* est un test du multiplicateur de Lagrange des hypothèses $R\varphi = r$, au même titre que les tests donnés par la formule ci-dessous, où I est remplacé par $\hat{I}$. Pour n'importe quelle matrice $\hat{I}\left(\hat{\varphi}_0\right)$ asymptotiquement équivalente à $I\left(\hat{\varphi}_0\right)$, $LM = \left(\left.\dfrac{\partial^2 L}{\partial\varphi\partial\varphi'}\right|_{\varphi=\hat{\varphi}_0}\right)' \hat{I}^{-1}\left(\hat{\varphi}_0\right)\left.\dfrac{\partial^2 L}{\partial\varphi\partial\varphi'}\right|_{\varphi=\hat{\varphi}_0} \xrightarrow{L} \chi^2_m$.

Asymptotiquement, $LR = W = LM$. Il n'en va pas de même pour un petit échantillon. Pour des hypothèses sur les coefficients d'un modèle linéaire, on montre que $W \geqslant LR \geqslant LM$. Donc toute hypothèse rejetée par le test *LM* est aussi rejetée par les tests *LR* et *W*.

5.2 Tests d'absence d'autocorrélation

Pour vérifier l'absence d'autocorrélation du terme d'erreur u, on dispose de plusieurs tests :

- **Test de Durbin et Watson** [DUR 1951] (voir section 9.2). Il ne peut s'appliquer lorsque les variables explicatives sont aléatoires [SPA 1986], en particulier lorsque la variable dépendante retardée fait partie des variables explicatives : dans ce cas, il est biaisé vers l'acceptation de l'hypothèse d'absence d'autocorrélation).

- **Test de Durbin** [DUR 1970]. Il est distribué asymptotiquement comme une $N(0, 1)$ sous l'hypothèse d'absence d'autocorrélation du terme d'erreur. Il est applicable quand la variable dépendante retardée fait partie des variables explicatives. Il se décline en deux versions : le « h normal » et le « h alternatif ». Le premier n'est pas toujours calculable (radicand négatif) et s'applique uniquement quand la variable dépendante a un seul retard. Le second s'applique même quand elle en a plusieurs. Quand elle n'en a qu'un, ces deux versions sont asymptotiquement équivalentes. Toutefois la puissance du « h alternatif » peut être très faible sur de petits échantillons [RAY 1994].
- **Test d'autocorrélation de Godfrey** [GOD(a) 1978], [GOD(b) 1978]. Il est applicable dans tous les cas, même quand la variable dépendante retardée fait partie des variables explicatives. Dans sa version LM, ce test est noté ici $LM^G(p)$ et est distribué asymptotiquement comme une χ^2_p, sous l'hypothèse H_0 que le terme d'erreur n'est pas autocorrélé (H_1 correspond à une autocorrélation d'ordre p, autorégressive ou en moyenne mobile). Dans sa version F, ce test est noté ici $F^G(p)$ et est distribué asymptotiquement comme une $F(p, n-k-p)$, sous l'hypothèse H_0. Les deux versions sont équivalentes asymptotiquement. Toutefois, sur de petits échantillons où elles ne sont qu'approximativement correctes, elles peuvent aboutir à des conclusions différentes. Ces tests sont pré-programmés dans Microfit par exemple.
- **Test d'autocorrélation de Breusch et Godfrey** [BRE 1981]. Il s'agit d'un test LM qui fonctionne comme le test de Godfrey et qui est distribué asymptotiquement comme une χ^2_p, sous l'hypothèse H_0 que le terme d'erreur n'est pas autocorrélé (H_1 correspond à une autocorrélation d'ordre p, autorégressive ou en moyenne mobile). Ce test est pré-programmé dans TSP par exemple.
- **Test $Q^*(p)$ de Ljung et Box** [LJU 1978]. Ce test « portmanteau » est distribué asymptotiquement comme une χ^2_p, sous l'hypothèse d'absence d'autocorrélation d'ordre p du terme d'erreur u. Sur de petits échantillons, il semble avoir de meilleures propriétés que le test $Q(p)$ de Box et Pierce [BOX 1970] dont il est une modification.

5.3 Tests d'homoscédasticité ou d'absence d'hétéroscédasticité

Pour vérifier l'absence d'hétéroscédasticité du terme d'erreur u, on dispose de plusieurs tests :

- **Test d'hétéroscédasticité de White** [WHI 1980]. Il est basé sur la régression du résidu au carré sur une constante et les produits croisés entre toutes les paires de variables explicatives différentes (constante comprise). On teste l'hypothèse nulle que tous les coefficients de cette régression sont nuls, sauf le terme constant. Sous cette hypothèse, le test est distribué selon une loi Chi-2 à $\big((k+1)k/2\big) - 1$ degrés de liberté.
- **Test *LMhet*.** Il vérifie l'hypothèse H_0 d'homoscédasticité (selon H_1, le carré du résidu dépend du carré de la valeur calculée $\hat{Y}_t$ de la variable dépendante). Sous H_0, *LMhet* est distribué asymptotiquement selon une χ^2_1.
- **Test $ARCH(p)$ d'Engle** [ENG 1982]. Il vérifie l'absence d'hétéroscédasticité autorégressive conditionnelle d'ordre p. Sous cette hypothèse nulle d'homoscédasticité, le test est distribué asymptotiquement selon une χ^2_p.
- **Test *LRhet*.** Il est basé sur une décomposition de l'échantillon global en deux sous-ensembles d'observations. Il vérifie l'hypothèse H_0 d'homoscédasticité du terme

d'erreur (selon H_1, la variance du terme d'erreur est différente dans chaque sous-échantillon). Sous l'hypothèse nulle, le test est distribué asymptotiquement selon une χ^2_1.

- **Test d'hétéroscédasticité de Breusch et Pagan** [BRE 1979]. Il fonctionne comme le test de White, en permettant d'introduire dans la régression des produits croisés faisant intervenir une liste arbitraire de variables.

5.4 Tests de normalité

Pour vérifier la normalité du terme d'erreur u, on dispose du **test de Jarque et Bera** [JAR 1987] [JAR 1981], qui est valable avec ou sans constante dans le modèle. Sous l'hypothèse nulle de normalité du terme d'erreur, ce test est distribué asymptotiquement comme une χ^2_2. On a tabulé également des valeurs critiques pour les petits échantillons.

5.5 Tests de stabilité des coefficients

Pour vérifier la constance dans le temps des coefficients β_i du modèle linéaire, on dispose par exemple des tests suivants :

- Premier **test de Chow** [CHO 1960]. Ce test est connu pour être un test de stabilité des coefficients de régression, ou encore un test d'analyse de covariance. Il sépare l'échantillon en deux ensembles de n_1 et n_2 observations ($n = n_1 + n_2$). Selon H_0, les coefficients β sont les mêmes sur les deux sous-échantillons, et selon H_1, ils sont différents, conditionnellement à l'hypothèse maintenue que la variance du terme d'erreur est la même. Il faut estimer le modèle sur l'ensemble des n observations, ce qui donne le vecteur de résidus e. Il faut aussi estimer le modèle sur les n_1 premières observations, ce qui donne le vecteur de résidus e_1, et ensuite sur les n_2 observations suivantes, ce qui donne le vecteur de résidus e_2. Sous H_0, on montre que

$$F^{CH} = \frac{(n_1 + n_2 - 2k)\,(e'e - e'_1 e_1 - e'_2 e_2)}{k\left(e'_1 e_1 + e'_2 e_2\right)} \sim F(k, n_1 + n_2 - 2k)$$

  Cette distribution est exacte pour de petits échantillons, valable à condition que le terme d'erreur soit normal. La version LM de ce test est la suivante : sous H_0, $kF^{CH} \xrightarrow{L} \chi^2_k$.

- Deuxième **test de Chow** [CHO 1960]. Ce test est connu sous le nom de test d'inaptitude à la prévision ou « predictive failure test » en anglais. Il s'emploie lorsque le nombre d'observations dans la deuxième partie de l'échantillon est inférieur au nombre de coefficients à estimer du modèle ($n_2 < k$). Il faut d'abord estimer le modèle sur l'ensemble des n observations, ce qui fournit le vecteur de résidus e_0 puis estimer le modèle sur les n_1 premières observations, ce qui fournit le vecteur de résidus e_1. Sous l'hypothèse H_0 d'égalité des coefficients β sur les deux sous-échantillons,

$$F^{CH} = \frac{(n_1 - k)\,(e'_0 e_0 - e'_1 e_1)}{n_2\left(e'_1 e_1\right)} \sim F(n_2, n_1 - k)$$

- La version LM de ce test est la suivante : sous H_0, $n_2 F^{CH} \xrightarrow{L} \chi^2_{n_2}$.

5.6 Tests de forme fonctionnelle adéquate

Le **test Reset** de Ramsey [RAM 1969] [RAM 1970] vérifie le caractère adéquat de l'hypothèse de linéarité en testant l'hypothèse nulle que le carré de la valeur estimée $\hat{Y}_t$ de la variable dépendante n'a pas d'influence additionnelle sur la variable dépendante Y_t. Ce test se décline en deux versions, et donc deux formules de calcul : une version *LM* et une version *F*. Sous l'hypothèse nulle, il est distribué asymptotiquement selon une loi χ^2_1 en version *LM*, et asymptotiquement selon une loi $F(1, n-k-1)$ en version *F*. On peut le généraliser en vérifiant l'effet de plusieurs puissances de $\hat{Y}_t$, jusqu'à une puissance p. Dans ce cas, le test *LM* est distribué asymptotiquement selon une loi χ^2_{p-1}, et le test *F* selon une loi $F\big(p-1, n-k-(p-1)\big)$, sous l'hypothèse nulle [ANS 1961].

Résumé

L'estimation d'un modèle linéaire par la technique des moindres carrés ordinaires est l'outil de base de l'économétrie. Une variable dépendante est modélisée comme une fonction d'une ou de plusieurs variables explicatives, à laquelle s'ajoute un terme d'erreur (une perturbation). Le choix de l'estimateur des moindres carrés ordinaires est adéquat lorsque le terme d'erreur du modèle est indépendant des variables explicatives. Cette condition garantit un estimateur MCO sans biais et convergent. Si le terme d'erreur n'est pas lié linéairement à ses valeurs passées, en d'autres termes s'il n'est pas autocorrélé et si sa variance est stable (on dit alors qu'il est « homoscédastique », par opposition à « hétéroscédastique »), c'est un bruit blanc. Sous cette condition, qui s'ajoute à celle d'indépendance, les estimateurs des moindres carrés ordinaires sont les plus précis des estimateurs linéaires sans biais. Il faut également que le terme d'erreur soit un bruit blanc pour que les formules utilisées par les logiciels pour calculer les écarts types des coefficients, donc leur « imprécision », soient correctes. La précision de l'estimation d'un coefficient est alors d'autant plus élevée que sa statistique t, qui est le rapport entre la valeur estimée et l'écart type, est élevée. Si le terme d'erreur est indépendant, de type bruit blanc et normal, on connaît la distribution sur de petits échantillons des estimateurs des moindres carrés ordinaires. On peut alors tester des hypothèses sur la valeur vraie des coefficients, ou calculer des intervalles de confiance pour ces coefficients, même avec peu d'observations. En l'absence de normalité du terme d'erreur, mais à condition qu'il soit indépendant des variables explicatives et qu'il soit un bruit blanc, on connaît quand même la distribution sur de grands échantillons. Elle permet de tester des hypothèses sur la valeur vraie des coefficients, ou de calculer des intervalles de confiance pour ces coefficients, quand il y a beaucoup d'observations. Comme il est indispensable de connaître les propriétés du terme d'erreur pour pouvoir interpréter les résultats d'estimation, on utilise des tests formels pour vérifier ces propriétés.



Problèmes et exercices

Exercice 1 Régression linéaire avec Excel

Énoncé

Estimez un modèle linéaire expliquant le taux de croissance de la consommation réelle en fonction du taux de croissance du revenu réel et du taux d'inflation. Les données de la consommation à prix constants, du revenu réel et de l'indice des prix, de 1960 à 1994, sont contenues dans le fichier USA.XLS téléchargeable sur le site Internet de l'éditeur (www.pearsoneducation.fr). Les définitions des variables principales de ce fichier sont les suivantes : CT est la consommation à prix constants, Y est le revenu disponible réel, P est l'indice des prix de la consommation. Ces estimations sont à réaliser au moyen du logiciel Excel.

Interprétez les résultats : les mouvements systématiques du taux de croissance de la consommation sont-ils expliqués par le modèle? Pouvez-vous déterminer si les coefficients du modèle sont estimés avec précision? Pouvez-vous tester les deux hypothèses suivantes : le taux d'inflation n'a pas d'influence sur la croissance de la consommation et l'élasticité de la consommation au revenu est égale à 1 à court terme?

Solution

Les logiciels tableurs comme Excel de Microsoft Office proposent quelques fonctions statistiques, dont l'estimation d'une régression linéaire par moindres carrés ordinaires. Comme ces logiciels sont très populaires, il est logique de commencer par montrer ce que l'on peut faire avec.

Le fichier de données fourni avec ce livre est un classeur Excel, nommé USA.XLS. Dans une feuille de calcul sont disposées des données sur plusieurs variables, pour les années 1960 à 1994. Chaque variable occupe une colonne, avec son nom en première ligne. La feuille de calcul se présente initialement comme à la figure 2.1, page suivante.

CT est la consommation totale en termes réels, R un taux d'intérêt à long terme, UR un taux de chômage, Y le revenu disponible réel, et P l'indice des prix.

On fait l'hypothèse que le taux de croissance de la consommation réelle est une fonction linéaire du taux de croissance du revenu disponible réel et du taux d'inflation. On suppose donc que :

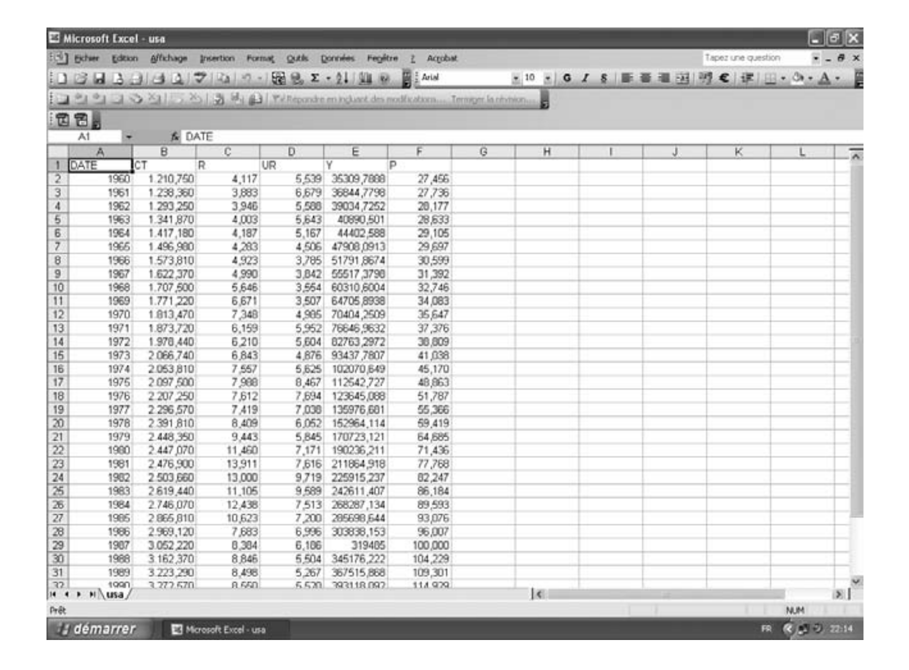
$$DLC_t = \beta_1 + \beta_2 DLY_t + \beta_3 DLP_t + u_t$$

$DLC_t = \ln(CT_t) - \ln(CT_{t-1})$ – taux de croissance de la consommation réelle.

$DLY_t = \ln(Y_t) - \ln(Y_{t-1})$ – taux de croissance du revenu réel.

$DLP_t = \ln(P_t) - \ln(P_{t-1})$ – taux de croissance de l'indice des prix.

Figure 2.1



DATE	CT	R	UR	Y	P
1960	1.210,750	4,117	5,539	35309,7888	27,456
1961	1.238,360	3,883	6,679	36844,7798	27,736
1962	1.293,250	3,946	5,588	39034,7252	28,177
1963	1.341,870	4,003	5,643	40890,501	28,633
1964	1.417,180	4,187	5,167	44402,588	29,105
1965	1.496,980	4,283	4,506	47908,0913	29,697
1966	1.573,810	4,923	3,785	51791,8674	30,599
1967	1.622,370	4,990	3,842	55517,3798	31,392
1968	1.707,500	5,646	3,554	60310,6004	32,746
1969	1.771,220	6,671	3,507	64705,8938	34,083
1970	1.813,470	7,348	4,985	70404,2509	35,647
1971	1.873,720	6,159	5,952	76646,9632	37,376
1972	1.978,440	6,210	5,604	82763,2972	38,809
1973	2.066,740	6,843	4,876	93437,7807	41,038
1974	2.053,810	7,557	5,625	102070,649	45,170
1975	2.097,500	7,988	8,467	112542,727	48,863
1976	2.207,250	7,612	7,694	123645,088	51,787
1977	2.296,570	7,419	7,038	135976,681	55,366
1978	2.391,810	8,409	6,052	152964,114	59,419
1979	2.448,350	9,443	5,845	170723,121	64,685
1980	2.447,070	11,460	7,171	190236,211	71,436
1981	2.476,900	13,911	7,616	211864,918	77,768
1982	2.503,660	13,000	9,719	225915,237	82,247
1983	2.619,440	11,105	9,589	242611,407	86,184
1984	2.746,070	12,438	7,513	268287,134	89,593
1985	2.865,810	10,623	7,200	285698,644	93,076
1986	2.969,120	7,683	6,996	303838,153	96,007
1987	3.052,220	8,384	6,186	319405	100,000
1988	3.162,370	8,846	5,504	345176,222	104,229
1989	3.223,290	8,498	5,267	367515,868	109,301
1990	3.272,570	8,650	5,520	393118,092	114,929

On veut estimer les paramètres β_1, β_2 et β_3 de ce modèle « vrai » supposé. On veut également vérifier dans quelle mesure ce modèle explique bien les fluctuations de la consommation.

Procédure à suivre

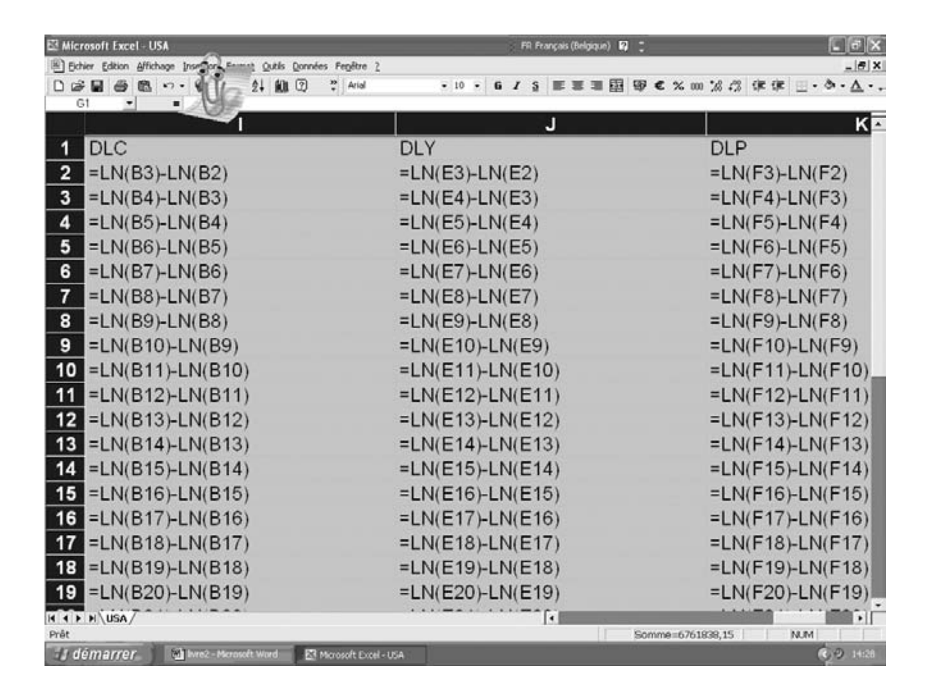
Il faut d'abord générer, dans la même feuille de calcul du classeur Excel, les transformations des données initiales dont vous avez besoin. Ces transformations sont les variables définies précédemment. Vous ne pouvez bien sûr les définir qu'à partir de 1961. En ce sens, recopiez les dates, à partir de 1961, dans une nouvelle colonne (H) et mettez dans les colonnes suivantes (I, J, K) les formules générant les nouvelles variables. À la première ligne, inscrivez les noms de ces nouvelles variables. Affichez la partie supérieure de ces colonnes avec les formules (voir figure 2.2, page ci-contre).

Imprimez la partie supérieure de cette feuille de calcul, qui comprend les valeurs de la figure 2.3, page ci-contre.

Pour estimer les paramètres du modèle « vrai » supposé par MCO, déroulez le menu Tools ou Outils, puis cliquez sur Data Analysis ou Utilitaire d'analyse. Un nouveau menu s'affiche. Cliquez sur Regression ou Régression linéaire. Une boîte de dialogue s'ouvre dans laquelle vous allez remplir ou cocher certaines cases.

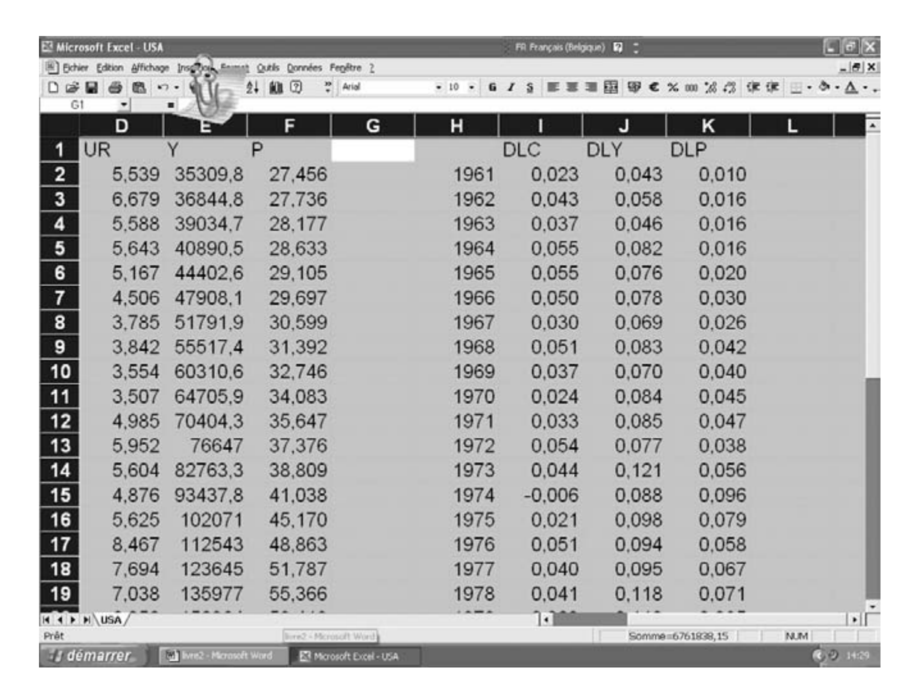
1. Dans la case Y-Range ou Plage pour la variable Y, placez la plage de cellules où se trouvent les informations (nom et données) de la variable dépendante, en l'occurrence I1:I35 (le logiciel écrit $I$1:$I$35).
2. Dans la case X-Range ou Plage pour les variables *X*, placez la plage de cellules où se trouvent les informations (noms et données) des variables explicatives, en l'occurrence J1:K35.
3. Cochez la case Labels ou Intitulé présent puisque les noms des variables sont inclus dans les plages de cellules en question.

Figure 2.2



	I	J	K
1	DLC	DLY	DLP
2	=LN(B3)-LN(B2)	=LN(E3)-LN(E2)	=LN(F3)-LN(F2)
3	=LN(B4)-LN(B3)	=LN(E4)-LN(E3)	=LN(F4)-LN(F3)
4	=LN(B5)-LN(B4)	=LN(E5)-LN(E4)	=LN(F5)-LN(F4)
5	=LN(B6)-LN(B5)	=LN(E6)-LN(E5)	=LN(F6)-LN(F5)
6	=LN(B7)-LN(B6)	=LN(E7)-LN(E6)	=LN(F7)-LN(F6)
7	=LN(B8)-LN(B7)	=LN(E8)-LN(E7)	=LN(F8)-LN(F7)
8	=LN(B9)-LN(B8)	=LN(E9)-LN(E8)	=LN(F9)-LN(F8)
9	=LN(B10)-LN(B9)	=LN(E10)-LN(E9)	=LN(F10)-LN(F9)
10	=LN(B11)-LN(B10)	=LN(E11)-LN(E10)	=LN(F11)-LN(F10)
11	=LN(B12)-LN(B11)	=LN(E12)-LN(E11)	=LN(F12)-LN(F11)
12	=LN(B13)-LN(B12)	=LN(E13)-LN(E12)	=LN(F13)-LN(F12)
13	=LN(B14)-LN(B13)	=LN(E14)-LN(E13)	=LN(F14)-LN(F13)
14	=LN(B15)-LN(B14)	=LN(E15)-LN(E14)	=LN(F15)-LN(F14)
15	=LN(B16)-LN(B15)	=LN(E16)-LN(E15)	=LN(F16)-LN(F15)
16	=LN(B17)-LN(B16)	=LN(E17)-LN(E16)	=LN(F17)-LN(F16)
17	=LN(B18)-LN(B17)	=LN(E18)-LN(E17)	=LN(F18)-LN(F17)
18	=LN(B19)-LN(B18)	=LN(E19)-LN(E18)	=LN(F19)-LN(F18)
19	=LN(B20)-LN(B19)	=LN(E20)-LN(E19)	=LN(F20)-LN(F19)

Figure 2.3

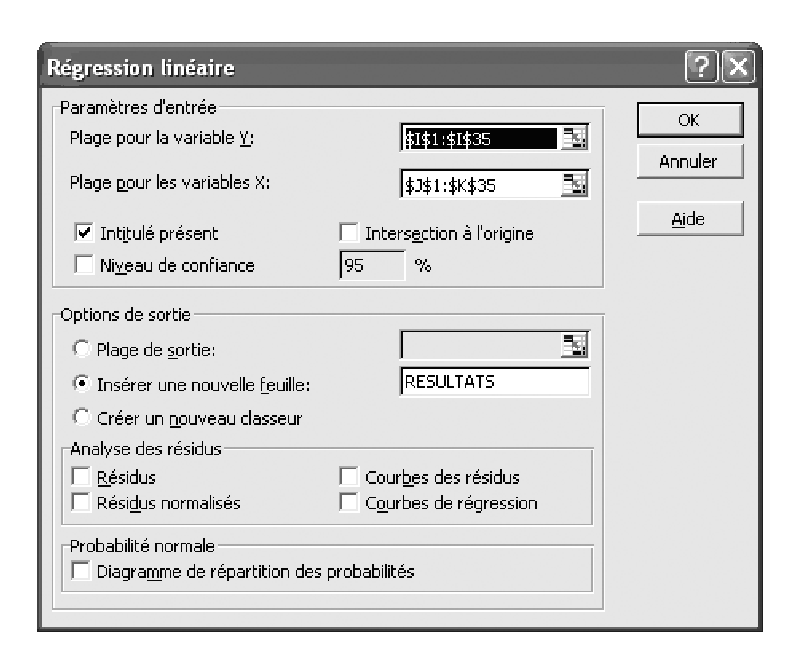


	D	E	F	G	H	I	J	K	L
1	UR	Y	P			DLC	DLY	DLP	
2	5,539	35309,8	27,456		1961	0,023	0,043	0,010	
3	6,679	36844,8	27,736		1962	0,043	0,058	0,016	
4	5,588	39034,7	28,177		1963	0,037	0,046	0,016	
5	5,643	40890,5	28,633		1964	0,055	0,082	0,016	
6	5,167	44402,6	29,105		1965	0,055	0,076	0,020	
7	4,506	47908,1	29,697		1966	0,050	0,078	0,030	
8	3,785	51791,9	30,599		1967	0,030	0,069	0,026	
9	3,842	55517,4	31,392		1968	0,051	0,083	0,042	
10	3,554	60310,6	32,746		1969	0,037	0,070	0,040	
11	3,507	64705,9	34,083		1970	0,024	0,084	0,045	
12	4,985	70404,3	35,647		1971	0,033	0,085	0,047	
13	5,952	76647	37,376		1972	0,054	0,077	0,038	
14	5,604	82763,3	38,809		1973	0,044	0,121	0,056	
15	4,876	93437,8	41,038		1974	-0,006	0,088	0,096	
16	5,625	102071	45,170		1975	0,021	0,098	0,079	
17	8,467	112543	48,863		1976	0,051	0,094	0,058	
18	7,694	123645	51,787		1977	0,040	0,095	0,067	
19	7,038	135977	55,366		1978	0,041	0,118	0,071	

4. Cochez la case Confidence Level ou Niveau de confiance si vous souhaitez obtenir une interprétation des tests de nullité de chaque coefficient vrai (c'est-à-dire le résultat de la comparaison de la statistique t à une distribution de Student à $n - k$, ici $34 - 3$, degrés de liberté) à un autre seuil de signification que celui par défaut (une marge de 95 % à gauche – et de 5 % à droite – signifie « 5 % de chances de rejeter l'hypothèse de nullité du paramètre alors qu'elle est vraie »). Ne cochez donc pas cette case si un test à 95 % vous satisfait.

5. Ne cochez pas la case Constant is 0 ou Intersection à l'origine si une constante figure dans le modèle « vrai » supposé, ce qui est le cas ici (β_1).
6. Cochez la case New Worksheet ou Insérer une nouvelle feuille pour que les résultats de l'estimation soient mis dans une nouvelle feuille de calcul.
7. Cochez la case Residuals ou Résidus pour obtenir les valeurs calculées de la variable dépendante et les résidus estimés.
8. Cochez la case Residual plots ou Courbes des résidus si vous voulez obtenir des graphiques du résidu en fonction de chaque variable explicative.
9. Cochez la case Line Fit Plot ou Courbe de régression si vous voulez obtenir un graphique des valeurs calculées de la variable dépendante en fonction de ses valeurs observées. Cliquez sur OK (voir figure 2.4).

Figure 2.4

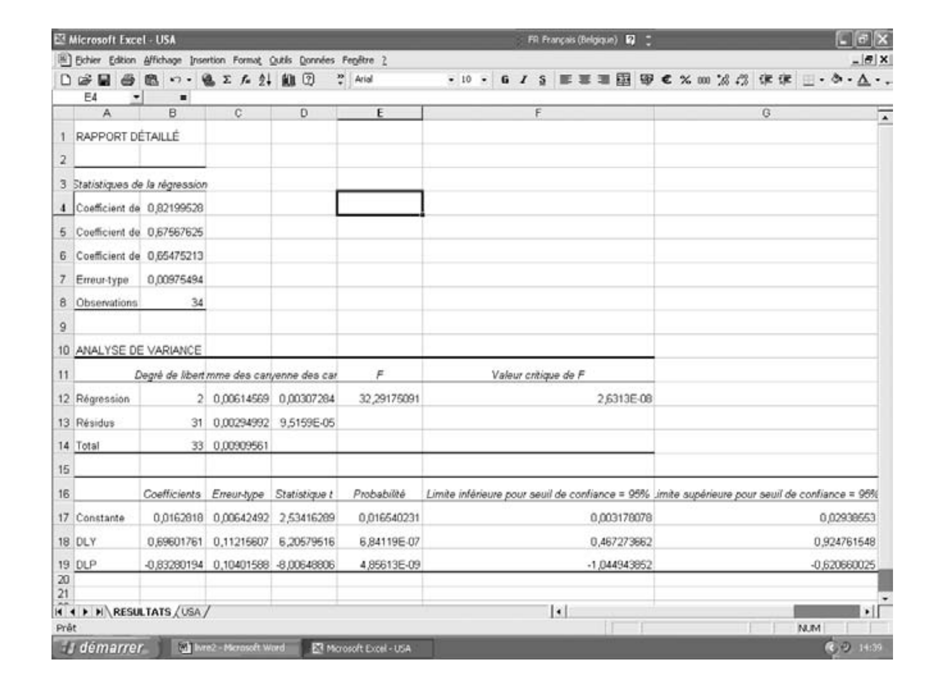


Vous obtenez les résultats de l'estimation par MCO dans une nouvelle feuille de calcul, RESULTATS, dans le même classeur (voir figure 2.5, page ci-contre).

Viennent d'abord des résultats descriptifs de la régression. La colonne Coefficients donne les valeurs des coefficients estimés par MCO (ici $\hat{\beta}_1^{MCO}$, $\hat{\beta}_2^{MCO}$ et $\hat{\beta}_3^{MCO}$). Le calcul de la formule de l'estimateur des moindres carrés ordinaires donne les résultats suivants : la valeur estimée $\hat{\beta}_1^{MCO}$ de la constante « vraie inconnue » β_1 vaut 0,016282 ; la valeur estimée $\hat{\beta}_2^{MCO}$ du coefficient « vrai inconnu » β_2 du taux de croissance du revenu DLY vaut 0,696020 ; la valeur estimée $\hat{\beta}_3^{MCO}$ du coefficient « vrai inconnu » β_3 du taux d'inflation DLP vaut −0,136783. Ces coefficients estimés sont plausibles économiquement. En effet, le taux de croissance de la consommation augmente lorsque le taux de croissance du revenu augmente, à taux d'inflation inchangé, et diminue lorsque le taux d'inflation augmente, à taux de croissance du revenu inchangé. C'est ce que suggère un raisonnement économique *a priori*.



Figure 2.5



	A	B	C	D	E	F	G
1	RAPPORT DÉTAILLÉ						
2							
3	*Statistiques de la régression*						
4	Coefficient de	0,82199528					
5	Coefficient de	0,67567625					
6	Coefficient de	0,65475213					
7	Erreur-type	0,00975494					
8	Observations	34					
9							
10	ANALYSE DE VARIANCE						
11		*Degré de libert*	*mme des car*	*yenne des car*	*F*	*Valeur critique de F*	
12	Régression	2	0,00614569	0,00307284	32,29175091	2,6313E-08	
13	Résidus	31	0,00294992	9,5159E-05			
14	Total	33	0,00909561				
15							
16		*Coefficients*	*Erreur-type*	*Statistique t*	*Probabilité*	*Limite inférieure pour seuil de confiance = 95%*	*Limite supérieure pour seuil de confiance = 95%*
17	Constante	0,0162818	0,00642492	2,53416289	0,016540231	0,003178078	0,02938553
18	DLY	0,69601761	0,11215607	6,20579616	6,84119E-07	0,467273662	0,924761548
19	DLP	-0,83280194	0,10401588	-8,00648806	4,85613E-09	-1,044943852	-0,620660025

La formule du calcul de ces coefficients estimés, utilisée par TSP, est bien sûr :

$$\hat{\beta}^{MCO} = \begin{pmatrix} \hat{\beta}_1^{MCO} \\ \hat{\beta}_2^{MCO} \\ \hat{\beta}_3^{MCO} \end{pmatrix} = (X'X)^{-1} X'Y$$

$$\text{où } X = \begin{pmatrix} 1 & DLY_{61} & DLP_{61} \\ 1 & DLY_{62} & DLP_{62} \\ \vdots & & \\ \vdots & & \\ 1 & DLY_{94} & DLP_{94} \end{pmatrix} \text{ et } Y = \begin{pmatrix} DLC_{61} \\ DLC_{62} \\ \vdots \\ \vdots \\ DLC_{94} \end{pmatrix}$$

Les valeurs R-squared et Adjusted R-squared sont le R^2 et le R^2 ajusté. Le R^2 est le pourcentage de variabilité de la variable dépendante DLC, qui est expliqué par la variabilité des variables explicatives DLY et DLP. La variabilité de DLY et de DLP explique 67 % de la variabilité de DLC.

Résultats disponibles

Les résultats suivants sont également disponibles (ils ne sont valables que si le terme d'erreur possède certaines propriétés spécifiques) :

- La colonne Erreur-type (Standard Errors en version anglaise) donne les valeurs des écarts types estimés par les estimateurs des coefficients. Ces valeurs ne sont valables que si le terme d'erreur vrai u est un bruit blanc.
- La colonne Statistiques t (t-stats en anglais) donne les valeurs des tests de nullité de la valeur vraie de chaque coefficient. En cas du nullité du coefficient vrai, la statistique t n'est distribuée selon une loi de Student à $n - k$ degrés de liberté que si le terme d'erreur

Exercices

vrai u est un bruit blanc et distribué normalement ; pour un grand échantillon, le caractère bruit blanc de u suffit (la normalité de u n'est pas nécessaire) et la statistique t est approximativement distribuée selon une $N(0, 1)$, qui est ce vers quoi tend une distribution Student pour beaucoup de degrés de liberté.

- La colonne Probabilité (p-value en version anglaise) permet d'interpréter la statistique t d'un coefficient : elle indique la probabilité qu'une réalisation de cette statistique soit supérieure à la valeur observée de ladite statistique, lorsque celle-ci est distribuée selon une loi de Student, donc lorsque le coefficient vrai est nul. Par exemple, si la valeur critique est inférieure à 0,05, on rejette l'hypothèse que ce coefficient est nul car, dans ce cas, la valeur réalisée de la statistique t se situe dans une zone de valeurs très peu probables ; vraisemblablement, il ne s'agit donc pas d'une loi de Student et le coefficient vrai n'est pas nul (s'il l'était, il s'agirait d'une loi de Student, à condition que u soit un bruit blanc et normal). L'information donnée par une p-value, pour tester la nullité des β_i, n'est évidemment valable que si la statistique t est distribuée selon une normale lorsque le coefficient vrai est nul, en d'autres termes que si u est un bruit blanc et (au moins pour un petit échantillon) normal.
- Les colonnes Limite inférieure pour seuil de confiance = 95 % (Lower 95 %) et Limite supérieure pour seuil de confiance = 95 % (Upper 95 %) donnent les bornes d'un intervalle de confiance pour les valeurs possibles de chaque coefficient. Elles ne sont valables que si u est un bruit blanc et normal (au moins pour un petit échantillon).
- La valeur Erreur-type (Standard Error) dans les Statistiques de la régression (Regression Statistics) est la valeur estimée de l'écart type vrai du terme d'erreur vrai u. Cette valeur n'est valable que si le terme d'erreur vrai u est un bruit blanc.
- La valeur F, dans le tableau ANALYSE DE VARIANCE (ANOVA), est un test de nullité simultanée de tous les coefficients, sauf de la constante. Lorsque cette hypothèse est vraie, la statistique F est distribuée selon une loi de Fischer à $k - 1$ et $n - k$ degrés de liberté, à condition toutefois que u soit un bruit blanc et normal.
- La valeur critique de F (Significance F) permet d'interpréter la statistique F : elle indique la probabilité qu'une réalisation de cette statistique soit supérieure à la valeur observée de ladite statistique, lorsque celle-ci est distribuée selon une loi de Fischer, donc lorsque tous les coefficients vrais, sauf la constante, sont nuls. Par exemple, si la valeur critique est inférieure à 0,05, on rejette l'hypothèse que tous les coefficients, sauf la constante, sont nuls car, dans ce cas, la valeur réalisée de la statistique F se situe dans une zone de valeurs très peu probables. Il ne s'agit vraisemblablement pas d'une loi de Fischer et les coefficients vrais ne sont pas tous nuls (s'ils l'étaient, il s'agirait d'une loi de Fischer, à condition que u soit un bruit blanc et normal). L'information donnée par la valeur critique de F, pour tester la nullité des β_i, n'est évidemment valable que si la statistique F est vraiment distribuée selon une loi de Fischer quand les coefficients vrais sont nuls, en d'autres termes que si u est un bruit blanc et normal.

Remarque

Sans analyser les résultats de tests d'autocorrélation, d'hétéroscédasticité, de normalité, de constance des coefficients, etc., il est impossible de savoir si les hypothèses de bruit blanc et de normalité du terme d'erreur sont vérifiées et par là même de déterminer la fiabilité des écarts types estimés des coefficients, des statistiques t, de leurs probabilités critiques, des bornes des intervalles de confiance, de l'écart type estimé du terme d'erreur, etc. La section suivante s'intéresse à de tels tests.

Exercice 2 Régression linéaire avec TSP™

Énoncé

Refaites l'exercice 1, mais cette fois en utilisant le logiciel spécialisé TSP (Times Series Processor). TSP est un produit de TSP International (**http://www.tspintl.com**).

Solution

Données

Placez les données à utiliser dans un fichier Excel appelé USA.XLS. Ce dernier doit être organisé afin de pouvoir être interprété par le logiciel TSP. En d'autres termes, il doit respecter les règles suivantes :

- Disposez les observations des différentes séries statistiques en colonnes. Chaque série temporelle occupe une colonne, avec en première ligne le nom donné à la variable.
- Si vous placez une colonne de dates, faites-la figurer en première colonne, en indiquant comme nom DATE.
- Enregistrez le fichier comme feuille de calcul (peu importe le choix de la version), et non comme classeur, dans le répertoire C :\ de votre ordinateur.

Le fichier USA.XLS contient des séries macroéconomiques pour les États-Unis, sur la période de 1960 à 1994. Les variables sont le suivantes :

CT est la consommation privée à prix constants.

Y est le revenu disponible réel.

P est l'indice des prix à la consommation.

Fonctionnement

Le logiciel TSP fonctionne sous Windows avec GiveWin. Vous pouvez l'utiliser en mode « batch » ou en mode « interactif » :

- En mode « batch », il faut d'abord écrire toutes les instructions successives dans un fichier. Vous avez le choix entre deux méthodes :
  - Soit vous utilisez l'éditeur WordPad accessible sous Windows. Attention : après la saisie, enregistrez le fichier avec l'extension .tsp. Ouvrez ensuite le logiciel GiveWin, cliquez sur File puis sur Open Text File, cliquez sur le nom du fichier .tsp puis sur Ouvrir (ou bien double-cliquez sur le fichier en question). Une boîte de dialogue présentant les instructions s'affiche.
  - Soit vous travaillez directement sous GiveWin. Pour cela, ouvrez GiveWin, cliquez sur File, sur New, puis sur Text et saisissez les instructions dans la boîte de dialogue. Enregistrez éventuellement le fichier (avec l'extension .tsp) en cliquant sur File puis sur Save as.

  Une fois les instructions saisies, toujours sous GiveWin, cliquez sur Modules puis sur Run TSP. Les résultats apparaissent dans une nouvelle fenêtre. Au besoin, enregistrez-les dans un fichier avec l'extension .out, en cliquant sur File puis sur Save as.
- En mode « interactif », ouvrez GiveWin, cliquez sur Modules, sur Add/Remove Module, puis sur Browse. Sélectionnez successivement Disk C :, Program Files, TSP4.5, TSPGW et cliquez sur Add. Cliquez de nouveau sur Modules, sur Start TSPGW, puis sur Start TSP Session et saisissez les instructions TSP, ligne par ligne, en appuyant sur Entrée à chaque fois. Les résultats s'affichent à l'écran après chaque commande lorsque celle-ci en produit.

Programme

```
FREQ A;
SMPL 60 94;
READ(FILE='C:\USA.XLS');
LC=LOG(CT);
LY=LOG(Y);
LP=LOG(P);
SMPL 61 94;
DLC=LC-LC(-1);
DLY=LY-LY(-1);
DLP=LP-LP(-1);
PLOT DLC;
PLOT DLY;
PLOT DLP;
PLOT DLC X DLY W;
GRAPH DLY DLC;
REGOPT(PVPRINT,STARS,QLAGS=2,LMLAGS=2) ALL;
NOPLOTS;
OLSQ DLC C DLY DLP;
```

`FREQ A;` informe que les données sont de fréquence annuelle (`A`). Pour des données mensuelles, utilisez `M`, pour des données trimestrielles `Q` ; pour d'autres fréquences ou des données en coupe instantanée, inscrivez `N`.

`SMPL 60 94;` informe que l'échantillon de données va de 1960 à 1994. Tant que cette instruction n'est pas remplacée dans le programme par une autre occurrence de `SMPL` avec d'autres dates, TSP réalise toutes les commandes qui suivent, sur la période 1960–1994.

`READ(FILE='C:\USA.XLS');` lit le fichier de données USA.XLS, qui se trouve dans le répertoire C : du disque dur.

Les instructions qui suivent génèrent, à partir des variables initiales figurant dans le fichier de données USA.XLS, d'autres variables qui sont des transformations des variables initiales.

`LC=LOG(CT);` génère le logarithme de la variable `CT`, pour toutes les dates de l'échantillon en cours. Cette nouvelle variable se nomme `LC`.

`LY=LOG(Y);` génère le logarithme de la variable `Y`, pour toutes les dates de l'échantillon en cours. Cette nouvelle variable se nomme `LY`.

`LP=LOG(P);` génère le logarithme de la variable `P`, pour toutes les dates de l'échantillon en cours. Cette nouvelle variable se nomme `LP`.

`SMPL 61 94;` modifie la période sur laquelle les instructions qui suivent doivent être réalisées. Celles-ci génèrent des variations des logarithmes des variables, et donc des taux de croissance. Ce calcul nécessite, pour chaque date, le logarithme de la variable à la période précédente : on ne peut donc commencer qu'en 1961.

`DLC=LC-LC(-1);` calcule la variation logarithmique, donc le taux de croissance, de la consommation. Cette nouvelle variable se nomme `DLC`.

`DLY=LY-LY(-1);` calcule la variation logarithmique, donc le taux de croissance, de `Y` (revenu disponible). Cette nouvelle variable se nomme `DLY`.

`DLP=LP-LP(-1);` calcule la variation logarithmique, donc le taux de croissance, de l'indice des prix à la consommation, soit le taux d'inflation. Cette nouvelle variable se nomme `DLP`.

`PLOT DLC`; produit le graphique de `DLC` en fonction du temps.

`PLOT DLY`; produit le graphique de `DLY` en fonction du temps.

`PLOT DLP`; produit le graphique de `DLP` en fonction du temps.

`PLOT DLC X DLY W`; produit le graphique de `DLC` et de `DLY` en fonction du temps. `DLC` est tracé sur l'axe des `X` et `DLY` sur l'axe des `W`.

`GRAPH DLY DLC`; produit le graphique de `DLC` en fonction de `DLY`.

Les instructions qui suivent ont pour but d'estimer le modèle « vrai » supposé :

$$DLC_t = \beta_1 + \beta_2 DLY_t + \beta_3 DLP_t + u_t$$

et donc de calculer par la formule des moindres carrés ordinaires le modèle estimé :

$$DLC_t = \hat{\beta}_1^{MCO} + \hat{\beta}_2^{MCO} DLY_t + \hat{\beta}_3^{MCO} DLP_t + e_t$$

Les coefficients estimés $\hat{\beta}_1^{MCO}$, $\hat{\beta}_2^{MCO}$ et $\hat{\beta}_3^{MCO}$ sont les valeurs estimées des coefficients « vrais inconnus » β_1, β_2 et β_3, à calculer à l'aide de la formule de l'estimateur de moindres carrées ordinaires.

`REGOPT(PVPRINT, STARS, QLAGS=2, LMLAGS=2) ALL`; demande à TSP de donner, à l'occasion des estimations fournies ultérieurement par l'instruction `OLSQ`, les résultats de tous (`ALL`) les tests. Grâce à `PVPRINT`, la probabilité critique (p-value) sera mentionnée à côté de chaque test. `STARS` informe sur l'éventuel rejet de l'hypothèse nulle correspondant au test. `QLAGS=2` exécutera le test « portmanteau » d'autocorrélation de Ljung et Box, pour un retard allant jusqu'à 2 périodes. `LMLAGS=2` exécutera le test LM d'autocorrélation de Breusch et Godfrey, également pour un retard allant jusqu'à 2 périodes.

`NOPLOTS`; demande à TSP de ne pas fournir de graphiques des résidus, etc, dans les résultats de l'instruction `OLSQ` qui suit.

`OLSQ DLC C DLY DLP`; estime par la formule des moindres carrés ordinaires une fonction de régression où `DLC` est la variable dépendante, `C` le nom réservé de la constante, `DLY` et `DLP` les variables explicatives.

Les principaux extraits des résultats sont les suivants :

```
Current sample: 1961 to 1994
                                  Equation 1
                                 ============

                    Method of estimation = Ordinary Least Squares

Dependent variable : DLC
Current sample : 1961 to 1994
Number of observations : 34

 Mean of dependent variable = .031883
Std. dev. of dependent var. = .016602
   Sum of squared residuals = .294985E-02
      Variance of residuals = .951563E-04
   Std. error of regression = .975481E-02
                  R-squared = .675683
         Adjusted R-squared = .654759
```

```
    Durbin-Watson statistic = 1.95346
Breusch/Godfrey LM : AR/MA1 = .111487E-02 [.973]
Breusch/Godfrey LM : AR/MA2 = .113874 [.945]
     Ljung-Box Q-statistic1 = .766268E-02 [.930]
     Ljung-Box Q-statistic2 = .314195 [.855]
  Wald nonlin. AR1 vs. lags = 1.40663 [.495]
                  ARCH test = .977662 [.323]
                 CuSum test = .628328 [.363]
               CuSumSq test = .106658 [1.00]
                  Chow test = .586376 [.629]
     LR het. test (w/ Chow) = -.934652 [1.00]
           White het.  test = 2.78845 [.733]
 Jarque-Bera normality test = .733369 [.693]
  F-statistic (zero slopes) = 32.2927 ** [.000]
   Akaike Information Crit. = -6.33802
 Schwarz Bayes. Info. Crit. = -9.04121
 Log of likelihood function = 110.746

           Estimated    Standard
Variable  Coefficient     Error        t-statistic    P-value
C           .016282     .642484E-02     2.53418    *   [.017]
DLY         .696020     .112155         6.20590    **  [.000]
DLP        -.136783     .082009        -1.66791        [.105]

             Variance Covariance of estimated coefficients

                         C             DLY            DLP
C             0.000041279
DLY          -0.00058796        0.012579
DLP          -0.00043960       0.0042425      0.0067254

               Correlation matrix of estimated coefficients

                         C             DLY             DLP
C                   1.0000
DLY                -0.81596         1.0000
DLP                -0.83433         0.46126         1.0000
```

Analyse des résultats de la régression linéaire : les résultats descriptifs de la régression

Il s'agit ici de résultats descriptifs de la régression, purement algébriques, dont la validité ne dépend pas de la véracité de certaines hypothèses de départ sur le modèle « vrai » supposé, et en particulier sur le terme d'erreur. Ils découlent mécaniquement de l'application de la formule des MCO aux séries de données, indépendamment de toute interprétation statistique.

Valeurs estimées des coefficients (Estimated Coefficient). La valeur estimée $\hat{\beta}_1^{MCO}$ de β_1 vaut 0,016282. La valeur estimée $\hat{\beta}_2^{MCO}$ de β_2 vaut 0,696020. La valeur estimée $\hat{\beta}_3^{MCO}$ de β_3 vaut $-0{,}136783$.



Moyenne empirique de la variable dépendante (Mean of Dependent Variable). C'est la moyenne des valeurs de la série DLC sur l'échantillon de 1961 à 1994 utilisé pour la régression : $\overline{DLC} = \frac{\sum_{t=61}^{94} DLC_t}{34}$. C'est une valeur estimée de l'espérance de cette variable.

Écart type empirique de la variable dépendante (Std. Dev. of Dependent Var.). C'est l'écart type estimé de la série DLC sur l'échantillon de 1961 à 1994 utilisé pour la régression : $S_{DLC} = \sqrt{\frac{\sum_{t=61}^{94} \left(DLC_t - \overline{DLC}\right)^2}{34-1}}$. C'est une valeur estimée de l'écart type, ou racine carrée de la variance, de cette variable.

Somme des carrés des résidus (Sum of Squared Residuals). C'est $e_{61}^2 + e_{62}^2 + \cdots + e_{94}^2$, soit la somme des carrés $e'e$ minimale à laquelle il est possible d'arriver.

Le R au carré (R-squared). C'est le pourcentage de variabilité de la variable dépendante DLC, en l'occurrence 67 %, qui est expliqué par la variabilité des variables explicatives DLY et DLP.

Le R au carré ajusté (Adjusted R-squared). C'est une variante du R^2, aménagée de manière à n'augmenter que si l'ajout d'une variable explicative supplémentaire induit une baisse « suffisante » de la somme des carrés des résidus (le R^2 traditionnel augmente systématiquement si l'on ajoute une variable explicative supplémentaire au modèle et qu'on le ré-estime).

Analyse des résultats de la régression linéaire : les tests de certaines hypothèses de départ sur le modèle vrai supposé

Il s'agit de résultats de tests de certaines hypothèses de départ sur le modèle « vrai » supposé, et en particulier sur le terme d'erreur. Ces hypothèses doivent être vérifiées pour que les résultats de la troisième catégorie soient corrects ou fiables.

On commence par vérifier l'hypothèse que le terme d'erreur vrai u est un bruit blanc, c'est-à-dire qu'il n'est ni autocorrélé, ni hétéroscédastique. Puis, on teste l'hypothèse de normalité, c'est-à-dire d'une distribution normale, du terme d'erreur vrai. Ensuite, on vérifie l'hypothèse de constance des coefficients « vrais inconnus » (ou hypothèse de stabilité paramétrique), c'est-à-dire que les coefficients « vrais inconnus » β_1, β_2 et β_3 sont vraiment les mêmes à toute période.

1) **Tests d'autocorrélation.** Il s'agit de vérifier l'hypothèse d'absence d'autocorrélation du terme d'erreur vrai u. Puisqu'il n'est pas observable, les différents tests disponibles pour vérifier cette hypothèse se fondent sur les résidus calculés e. La tâche n'est pas facile, car même en l'absence d'autocorrélation de u, e est quand même autocorrélé [(1)]. La détermination des formules de bons tests d'absence d'autocorrélation de u est donc un exercice sophistiqué pour les mathématiciens statisticiens. Le logiciel TSP calcule plusieurs de ces tests.
   - **Test du Durbin et Watson.** Il n'est valable que si la variable dépendante retardée n'est pas incluse dans les variables explicatives, ce qui est le cas dans cette application (DLC_{t-1} ne figure pas parmi les variables explicatives).

1. En effet $e = Mu$. Si u est un bruit blanc, ce qui implique entre autres l'absence d'autocorrélation, $\Sigma_u = \sigma_u^2 I_n$. Donc la matrice de variance et de covariance de e est $\Sigma_e = M\Sigma_u M' = M\sigma_u^2 I_n M' = \sigma_u^2 M I_n M' = \sigma_u^2 MM'\sigma_u^2 M$ qui n'est pas une matrice diagonale : e est donc autocorrélé.

Dans une table du test de Durbin et Watson, on trouve, en fonction des valeurs de k (ici 3) et de n (ici 34) des valeurs critiques dL et dU, pour un seuil critique de 5 % ou de 2,5 %, au choix de l'analyste. On calcule alors $4 - dL$ et $4 - dU$ et on interprète le test. Si la valeur calculée du test est inférieure à dL, on adopte l'hypothèse d'une autocorrélation positive du terme d'erreur vrai u. Si elle est située entre dL et dU, ou entre $4 - dU$ et $4 - dL$, on ne peut se prononcer ; si elle est supérieure à $4 - dL$, on choisit l'hypothèse d'une autocorrélation négative ; si elle est située entre dU et $4 - dU$, on ne rejette pas l'hypothèse d'absence d'autocorrélation du terme d'erreur vrai u.

Pour $n = 34$ et $k = 3$, les tables de Durbin et Watson indiquent les valeurs critiques $dL = 1{,}33$ et $dU = 1{,}48$. On en déduit l'absence d'autocorrélation de u.

- **Tests *Q* de Ljung et Box jusqu'à 2 retards.** Un test $Q(i)$ de Ljung et Box vérifie l'hypothèse nulle H_0 que le terme d'erreur vrai u_t n'est lié linéairement à aucune de ses i valeurs précédentes : $\text{Cov}(u_t, u_{t-1}) = 0$, $\text{Cov}(u_t, u_{t-2}) = 0 \ldots \text{Cov}(u_t, u_{t-i}) = 0$.

  Le logiciel calcule ici les tests $Q(1)$ et $Q(2)$. L'hypothèse nulle H_0 de $Q(1)$ est l'absence d'autocorrélation du premier ordre de u ($\text{Cov}(u_t, u_{t-1}) = 0$) et l'hypothèse nulle H_0 de $Q(2)$ est l'absence d'autocorrélation du premier ordre *et* du second ordre de u ($\text{Cov}(u_t, u_{t-1}) = 0$ et $\text{Cov}(u_t, u_{t-2}) = 0$).

  Sous l'hypothèse nulle d'autocovariances nulles ou autocorrélations nulles (ou absence d'autocorrélation du terme d'erreur), le test $Q(i)$ est distribué comme une Chi-2 à i degrés de liberté. On rejette l'hypothèse nulle si la valeur observée de $Q(i)$ est supérieure à la valeur critique d'une table de Chi-2 à i degrés de liberté, pour un seuil de signification donné (par exemple 5 %), en d'autres termes si la probabilité critique du test $Q(i)$ est inférieure au seuil critique choisi (par exemple 0,05). Ici, on ne rejette pas l'hypothèse d'absence d'autocorrélation du terme d'erreur puisque $0{,}930 \gg 0{,}05$ et $0{,}855 \gg 0{,}05$.

- **Test *LM* de Breusch Godfrey jusqu'à 2 retards.** Un test $LM^{aut}(i)$ de Breusch Godfrey vérifie l'hypothèse nulle H_0 que le terme d'erreur vrai u_t n'est lié linéairement à aucune de ses i valeurs précédentes : $\text{Cov}(u_t, u_{t-1}) = 0$, $\text{Cov}(u_t, u_{t-2}) = 0 \ldots \text{Cov}(u_t, u_{t-i}) = 0$.

  Le logiciel calcule ici les tests $LM^{aut}(1)$ et $LM^{aut}(2)$. L'hypothèse nulle H_0 de $LM^{aut}(1)$ est l'absence d'autocorrélation du premier ordre de u ($\text{Cov}(u_t, u_{t-1}) = 0$) et l'hypothèse nulle H_0 de $LM^{aut}(2)$ est l'absence d'autocorrélation du premier ordre *et* du second ordre de u ($\text{Cov}(u_t, u_{t-1}) = 0$ et $\text{Cov}(u_t, u_{t-2}) = 0$).

  Sous l'hypothèse nulle d'autocovariances nulles ou autocorrélation nulles (ou absence d'autocorrélation du terme d'erreur), le test $LM^{aut}(i)$ est distribué comme une Chi-2 à $i+k-1$ degrés de liberté. On rejette l'hypothèse nulle si la valeur observée de $LM^{aut}(i)$ est supérieure à la valeur critique d'une table de Chi-2 à $i + k - 1$ degrés de liberté, pour un seuil de signification donné (par exemple 5 %), en d'autres termes si la probabilité critique du test $LM^{aut}(i)$ est inférieure au seuil critique choisi (par exemple 0,05). Ici, on ne rejette pas l'hypothèse d'absence d'autocorrélation du terme d'erreur puisque $0{,}973 > 0{,}05$ et $0{,}945 > 0{,}05$

2) **Tests d'hétéroscédasticité.** Il s'agit de tester l'hypothèse d'absence d'hétéroscédasticité du terme d'erreur u, c'est-à-dire que la variance du terme d'erreur vrai est la même à toute période.

   - **Test de White.** Sous l'hypothèse nulle d'absence d'hétéroscédasticité (dépendance de la variance de u aux valeurs des variables explicatives), le test White het est distribué

selon une Chi-2 à $((k+1)k/2) - 1$ degrés de liberté. On rejette l'hypothèse nulle si la valeur observée du test de White het est supérieure à la valeur critique d'une table de Chi-2 à $((k+1)k/2) - 1$ degrés de liberté, pour un seuil de signification donné (par exemple 5 %), en d'autres termes si la probabilité critique du test White het est inférieure au seuil critique choisi (par exemple 0,05). Ici on ne rejette pas l'hypothèse d'absence d'hétéroscédasticité puisque $0{,}733 > 0{,}05$.

- **Test LR het w/Chow.** Sous l'hypothèse nulle d'absence d'hétéroscédasticité (égalité de la variance de u sur deux sous-parties de l'échantillon), le test LR het w/Chow est distribué selon une Chi-2 à 1 degré de liberté. On rejette l'hypothèse nulle si la valeur observée du test LR het w/Chow est supérieure à la valeur critique d'une table de Chi-2 à 1 degré de liberté, pour un seuil de signification donné (par exemple 5 %), en d'autres termes si la probabilité critique du test LR het w/Chow est inférieure au seuil critique choisi (par exemple 0,05). Ici on ne rejette pas l'hypothèse d'absence d'hétéroscédasticité puisque $1{,}00 > 0{,}05$.
- **Test ARCH.** Sous l'hypothèse nulle d'absence d'hétéroscédasticité autorégressive du premier ordre (voir chapitre 7 sur les modèles ARCH et GARCH), le test LR het ARCH est distribué selon une Chi-2 à 1 degré de liberté. On rejette l'hypothèse nulle si la valeur observée du test ARCH est supérieure à la valeur critique d'une table de Chi-2 à 1 degré de liberté, pour un seuil de signification donné (par exemple 5 %), en d'autres termes si la probabilité critique du test ARCH est inférieure au seuil critique choisi (par exemple 0,05). Ici on ne rejette pas l'hypothèse d'absence d' hétéroscédasticité autorégressive puisque $0{,}323 > 0{,}05$.

3) **Tests de normalité.** On vérifie ensuite l'hypothèse de normalité du terme d'erreur vrai u.

- **Test de normalité de Jarque et Bera.** Il vérifie l'hypothèse nulle de normalité de u : $u \sim N(.,.)$. Sous H_0, le test de Jarque et Bera est distribué selon une Chi-2 à 2 degrés de liberté. On rejette l'hypothèse nulle si la valeur observée de Jarque et Bera est supérieure à la valeur critique d'une table de Chi-2 à 2 degrés de liberté, pour un seuil de signification donné (par exemple 5 %), en d'autres termes si la probabilité critique du test de Jarque et Bera est inférieure au seuil critique choisi (par exemple 0,05). Ici on ne rejette pas l'hypothèse de normalité de u puisque $0{,}693 > 0{,}05$.

4) **Tests de constance des coefficients.** On vérifie ensuite l'hypothèse de constance des coefficients « vrais inconnus ». On vérifie donc l'hypothèse que les coefficients vrais inconnus β_1, β_2 et β_3 sont vraiment les mêmes en toute période.

- **Test de Chow.** Sous l'hypothèse nulle de stabilité des coefficients (égalité des coefficients sur deux sous-parties de l'échantillon), le test de Chow est distribué selon une Fischer à k degrés de liberté au numérateur et $n - 2k$ degrés de liberté au dénominateur. On rejette H_0 si la valeur observée de Chow est supérieure à la valeur critique d'une table de Chi-2 à 1 degré de liberté, pour un seuil de signification donné (par exemple 5 %), en d'autres termes si la probabilité critique du test de Chow est inférieure au seuil critique choisi (par exemple 0,05). Ici, on ne rejette pas l'hypothèse de stabilité des coefficients puisque $0{,}629 > 0{,}05$.
- **Test Cusum** (Cumulated Sum of Residuals). Sous l'hypothèse nulle de stabilité des coefficients (égalité des coefficients sur deux sous-parties de l'échantillon), le test Cusum est distribué selon une distribution spéciale de Durbin [DUR 1971]. On rejette H_0 si la valeur observée du test Cusum est supérieure à la valeur critique d'une table de la distribution spéciale, pour un seuil de signification donné (par exemple 5 %), en d'autres termes si la probabilité critique du test Cusum est inférieure au

seuil critique choisi (par exemple 0,05). Ici, on ne rejette pas l'hypothèse de stabilité des coefficients puisque 0,363 > 0,05.

- **Test Cusumsq** (Cumulated Sum of Squared Residuals). Sous l'hypothèse nulle de stabilité des coefficients (égalité des coefficients sur deux sous-parties de l'échantillon), le test Cusumsq est distribué selon une distribution spéciale de Durbin [DUR 1969]. On rejette H_0 si la valeur observée du test Cusumsq est supérieure à la valeur critique d'une table de distribution spéciale de Durbin de 1969 [DUR 1969] pour un seuil de signification donné (par exemple 5 %), en d'autres termes si la probabilité critique du test Cusumsq est inférieure au seuil critique choisi (par exemple 0,05). Ici on ne rejette pas l'hypothèse de stabilité des coefficients puisque 1,000 > 0,05

Analyse des résultats de la régression linéaire : les résultats qui ne sont corrects ou fiables que sous certaines hypothèses sur le modèle vrai supposé

Il s'agit de résultats corrects ou fiables dont la validité est conditionnelle à la véracité de certaines hypothèses de départ sur le modèle « vrai » supposé, et en particulier sur le terme d'erreur. Ces hypothèses ont été vérifiées à travers les résultats de la deuxième catégorie.

Écarts types estimés des coefficients estimés. Il s'agit des Standard Errors fournis par le logiciel TSP à côté des coefficients estimés (Estimated Coefficients). Ce sont les valeurs estimées des écarts types « vrais inconnus » des estimateurs $\hat{\beta}_1^{MCO}$ des coefficients « vrais inconnus » β_i. Ce sont les éléments diagonaux de la matrice de variance et de covariance estimée des coefficients estimés, calculée par la formule :

$$\hat{\Sigma}_{\hat{\beta}^{MCO}} = \left(\frac{e'e}{34-3}\right)(X'X)^{-1}$$

Cette formule n'est valable que si le terme d'erreur est un bruit blanc, donc s'il n'est ni autocorrélé ni hétéroscédastique. C'est en effet sous ces hypothèses que la formule de $\hat{\Sigma}_{\hat{\beta}^{MCO}}$, utilisée par les logiciels, est obtenue mathématiquement.

Matrice de variance et de covariance estimée des coefficients estimés. Toutes les valeurs de la matrice sont présentées dans le tableau Variance Covariance of Estimated Coefficients. Il n'y figure que la diagonale principale et le triangle inférieur gauche de la matrice parce que celle-ci est symétrique (son triangle supérieur droit est la transposée de son triangle inférieur gauche).

Écart type estimé du terme d'erreur vrai. Il s'agit du résultat Standard Error of Regression fourni par le logiciel. C'est la valeur estimée $\hat{\sigma}_u$ de σ_u, calculée d'après la formule de l'estimateur sans biais $\hat{\sigma}_u^2$ de σ_u^2 : $\hat{\sigma}_u^2 = \dfrac{e'e}{34-3}$. Cette formule n'est valable que si le terme d'erreur est un bruit blanc, donc s'il n'est ni autocorrélé ni hétéroscédastique. C'est en effet sous ces hypothèses que la formule de $\hat{\sigma}_u^2$ est obtenue mathématiquement.

Statistiques *t*. Les *t*-stats sont le rapport entre le coefficient estimé et l'écart type estimé de l'estimateur du coefficient. En termes des résultats de TSP, chaque statistique *t* est donc le rapport entre l'Estimated Coefficient et le Standard Error.

$$t\text{-}stat_i = \frac{\hat{\beta}_i^{MCO}}{S_{\hat{\beta}_i^{MCO}}}$$

Très utilisé, la statistique *t* d'un coefficient permet de vérifier l'hypothèse, importante, que sa valeur « vraie inconnue » β_i est nulle. Mais ce test n'est valable que si le terme

d'erreur est vraiment un bruit blanc (ni autocorrélé ni hétéroscédastique), et si l'un des deux points suivants est vrai :

- Soit le terme d'erreur vrai u est réellement normal, auquel cas on peut tester l'hypothèse pour n'importe quelle taille d'échantillon (même si elle est petite), en comparant la valeur du test à la valeur critique d'une t_{n-k} à un seuil de signification donné (le plus souvent 5 %, 1 % ou 10 %).
- Soit le terme d'erreur vrai u n'est pas distribué selon une normale, mais l'échantillon est grand ; dans ce cas, on peut tester approximativement l'hypothèse, en comparant la valeur du test à la valeur critique d'une $N(0, 1)$.

Le test de la statistique t s'utilise ainsi : on sait que pour la valeur « vraie inconnue » β_i, $\dfrac{\hat{\beta}_i^{MCO} - \beta_i}{S_{\hat{\beta}_i^{MCO}}}$ est distribué selon une loi de Student t_{n-k} (propriété statistique 8 des estimateurs de MCO si u est indépendant de X, de type bruit blanc et normal). Donc, si la valeur vraie $\beta_i = 0$, $t\text{-}stat_i$ (la statistique t de β_i) est distribuée selon une loi de Student t_{n-k}. Pour vérifier que $\beta_i = 0$, il suffit donc de s'assurer de la distribution de la statistique t de β_i. Pour cela, on regarde si l'unique réalisation que l'on en a, fournie par le logiciel (colonne *t-stat* ou *t-ratio*), a une valeur acceptable pour une variable distribuée selon une loi de Student t_{n-k}. On trouve dans les tables de la Student t_{n-k} une valeur critique $vc_{5\,\%}$; cela signifie qu'une variable t_{n-k} n'a que 5 chances sur 100 de se réaliser à une valeur supérieure à $vc_{5\,\%}$. Si la valeur absolue de la valeur réalisée observée de la $t\text{-}stat_i$ est supérieure à $vc_{5\,\%}$, on en déduit que la $t\text{-}stat_i$ n'est vraisemblablement pas distribuée selon une t_{n-k} (puisqu'elle s'est réalisée dans une zone de valeurs peu probables pour une loi de Student t_{n-k}) et que β_i n'est vraisemblablement pas égal à 0 (s'il l'était, $t\text{-}stat_i$ serait forcément distribuée par une t_{n-k}). Donc, si la valeur absolue de $t_{n-k>}$ est supérieure à $vc_{5\,\%}$, on rejette l'hypothèse H_0 que $\beta_i = 0$. Si la valeur absolue de t_{n-k} est inférieure à $vc_{5\,\%}$, on ne la rejette pas.

Il faut donc disposer d'une table de distribution Student pour faire ce type de test. La plupart des logiciels ont ces tables en mémoire et peuvent interpréter le test pour l'utilisateur. Au lieu de donner la valeur critique issue de la table de t_{n-k}, les logiciels comme TSP, E-VIEWS, Easyreg et Microfit donnent la probabilité critique de la valeur observée du test $t\text{-}stat_i$, c'est-à-dire la probabilité qu'une variable t_{n-k} ait une réalisation supérieure à la valeur observée du test $t\text{-}stat_i$ si celle-ci est supérieure à 0, ou inférieure à la valeur observée du test $t\text{-}stat_i$ si celle-ci est inférieure à 0. Si cette probabilité critique est inférieure (respectivement supérieure) à 0,05, cela implique forcément que la valeur absolue de la valeur observée du test $t\text{-}stat_i$ est supérieure (respectivement inférieure) à la valeur critique $vc_{5\,\%}$; on rejette alors l'hypothèse nulle (respectivement, on ne la rejette pas). Ici, on rejette l'hypothèse $\beta_2 = 0$, mais non les hypothèses $\beta_1 = 0$ et $\beta_3 = 0$.

Bien sûr, la fiabilité du test n'est pas parfaite : β_i étant vraiment égal à 0 et la $t\text{-}stat_i$ étant donc vraiment distribuée selon une t_{n-k}, sa valeur réalisée a quand même 5 chances sur 100 d'être supérieure à $vc_{0,05}$. En pareil cas, on rejette l'hypothèse alors qu'elle est vraie : c'est l'erreur de première espèce. En utilisant un seuil de signification de 5 %, la probabilité de faire cette erreur est de 0,05. On peut diminuer ce risque en choisissant plutôt la valeur critique $vc_{0,01}$: on n'a plus que 1 chance sur 100 de rejeter l'hypothèse nulle alors qu'elle est vraie. D'un autre côté, quand on rejette moins souvent l'hypothèse, on augmente la probabilité de l'autre type d'erreur possible : celle de ne pas rejeter l'hypothèse alors qu'elle est fausse !

Par ailleurs, si l'on sait précisément, une fois le seuil critique choisi, quelle est la probabilité du type d'erreur de première espèce (rejet d'une hypothèse vraie), on ne sait généralement pas quelle est la probabilité de l'autre type d'erreur (« non-rejet » d'une hypothèse fausse).

On peut donc dire « On rejette l'hypothèse H_0 avec 95 chances sur 100 de ne pas se tromper », mais il faut absolument éviter des phrases du type de « On ne rejette pas (ou on accepte) l'hypothèse H_0 avec 95 chances sur 100 de ne pas se tromper » qui n'ont aucun sens. Tout ce que l'on sait en effet, c'est que l'on a 95 chances sur 100 de ne pas faire l'erreur de première espèce, mais l'on ne sait généralement pas quelle est la probabilité de faire l'erreur de deuxième espèce. Quand on ne rejette pas une hypothèse (donc quand on l'accepte), on ne sait pas quel est le pourcentage de chances de ne pas se tromper.

Important

Si on ne rejette pas l'hypothèse que $\beta_i = 0$, cela ne veut pas dire qu'on est certain que $\beta_i = 0$. Cela signifie simplement que 0 fait partie des valeurs vraies possibles de β_i parmi toutes les autres valeurs qui ne sont pas rejetées par un test d'hypothèse similaire (voir plus loin). Quand on n'écarte pas l'hypothèse que $\beta_i = 0$, on peut vérifier que d'autres valeurs positives et négatives, autour de 0, ne sont pas rejetées non plus par un test d'hypothèse. En fait, avant l'estimation d'une équation, on ignore la valeur des coefficients « vrais inconnus », alors qu'après, on connaît un ensemble réduit de valeurs qui ne sont pas rejetées par un test d'hypothèse (à un seuil critique arbitrairement fixé) : c'est l'intervalle de confiance. En d'autres termes, quand on n'écarte pas l'hypothèse que $\beta_i = 0$, l'analyse permet de réduire faiblement l'ignorance initiale sur la valeur vraie du coefficient β_i puisque l'on ne sait toujours pas s'il est nul, négatif ou positif.

L'élasticité de la consommation au revenu est-elle égale à 1 à court terme?

Cette question revient à demander si β_2 est égal à 1. On sait que, pour la valeur « vraie inconnue » de β_i, $\frac{\hat{\beta}_i^{MCO} - \beta_i}{S_{\hat{\beta}_i^{MCO}}} \sim t_{n-k}$ si le terme d'erreur u est un bruit blanc, indépendant de X et distribué normalement. Dans le cas où le terme d'erreur u a une autre distribution, mais est quand même un bruit blanc, indépendant de X, on sait que pour une grande taille n d'échantillon : $\frac{\hat{\beta}_i^{MCO} - \beta_i}{S_{\hat{\beta}_i^{MCO}}} \sim N(0,1)$. Ces résultats permettent de tester toutes les hypothèses souhaitées sur la valeur vraie d'un coefficient β_i, pour n'importe quel $i = 1 \dots k$. On peut tester par exemple l'hypothèse que $\beta_i = 1$. On sait que si elle est vraie (si $\beta_i = 1$), $\frac{\hat{\beta}_i^{MCO} - 1}{S_{\hat{\beta}_i^{MCO}}} \sim t_{n-k}$. Pour vérifier ce qu'il en est, il suffit donc de s'assurer que $\frac{\hat{\beta}_i^{MCO} - 1}{S_{\hat{\beta}_i^{MCO}}} \sim t_{n-k}$. On trouve dans une table de t_{n-k} une valeur critique $vc_{5\,\%}$ telle qu'une variable aléatoire t_{n-k} a 95 chances sur 100 d'être inférieure à $vc_{5\,\%}$ et seulement 5 chances sur 100 de la dépasser. Si la valeur calculée de $\left|\frac{\hat{\beta}_i^{MCO} - 1}{S_{\hat{\beta}_i^{MCO}}}\right|$ est supérieure à la valeur critique $vc_{5\,\%}$ (de la table de la distribution t_{n-k}) à un seuil de signification de 5 % (par exemple), on rejette l'hypothèse que $\frac{\hat{\beta}_i^{MCO} - 1}{S_{\hat{\beta}_i^{MCO}}} \sim t_{n-k}$ et donc que $\beta_i = 1$. On pense

en effet que $\left|\frac{\hat{\beta}_i^{MCO}-1}{S_{\hat{\beta}_i^{MCO}}}\right|$ s'est réalisé dans une zone peu probable (seulement 5 chances sur 100) pour une distribution t_{n-k} et donc qu'il n'est pas vraisemblable pas que $\beta_i = 1$; l'hypothèse $\beta_i = 1$ est vraisemblablement fausse.

Encore une fois, si l'on ne rejette pas l'hypothèse que $\beta_i = 1$, cela ne veut pas dire qu'on est certain que $\beta_i = 1$. Cela signifie simplement que 1 fait partie des valeurs vraies possibles de β_i parmi toutes les autres valeurs qui ne sont pas rejetées par le test d'hypothèse.

Ici, on rejette l'hypothèse $\beta_2 = 1$ car $\left|\frac{\hat{\beta}_2^{MCO}-1}{S_{\hat{\beta}_2^{MCO}}}\right| = \left|\frac{0{,}696020-1}{0{,}112155}\right| = 2{,}71$, qui est une valeur supérieure à la valeur critique à 5 % d'une distribution t_{34-3}.

Au lieu d'utiliser manuellement les tables, on peut automatiser la procédure pour que TSP vérifie l'hypothèse en calculant la probabilité critique de la valeur calculée du test précédent. Pour cela, il faut ajouter les instructions suivantes au programme, juste après l'instruction `OLSQ` :

```
SET TTEST=(@COEF(2)-1)/@SES(2);
SET DFT=@NOB-@NCID;
CDF(T,DF=DFT) TTEST;
```

F-statistic (zero slopes). Il s'agit du test de l'hypothèse $\beta_2 = \beta_3 = 0$. On rejette cette hypothèse au seuil de 5 % parce que la probabilité critique associée (entre [ ]) est inférieure à 0,05.

Exercice 3 Régression linéaire avec SPSS

Énoncé

Refaites l'exercice 1 en utilisant le logiciel SPSS (le site Internet de SPSS France est **http://www.spss.com/fr/index.htm**).

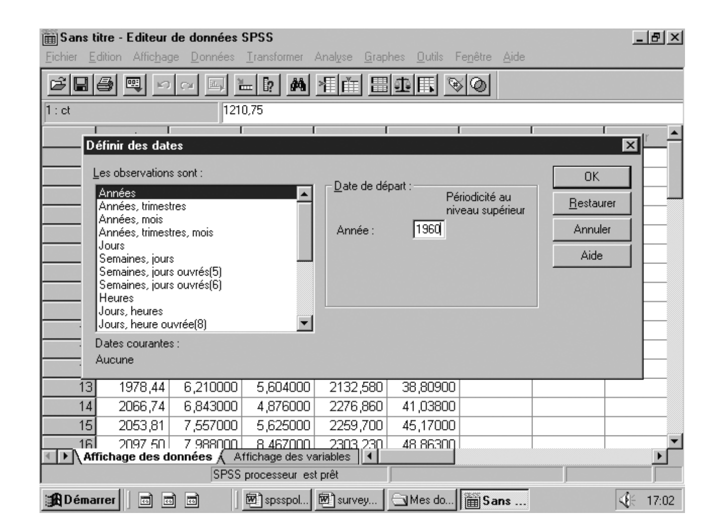
Solution

Éditez le fichier USA.XLS original (c'est-à-dire simplement les données brutes, avant toute manipulation effectuée sous Excel) sous Excel et supprimez la colonne de dates. Le fichier de données se présente donc comme à la figure 2.1, page 46, mais sans les dates ; les colonnes A, B, C, D, E sont maintenant occupées par les variables CT, R, UR, Y, P.

Quittez Excel. Ouvrez SPSS. Cliquez successivement sur Fichier/Ouvrir/Données. Une boîte de dialogue s'ouvre ; choisissez Mes Documents dans Rechercher dans, et Excel dans Type de fichier, puis cliquez sur usa.xls pour le sélectionner. Cliquez sur Ouvrir. La boîte de dialogue Ouvrir la source de données Excel s'ouvre. Cochez Lire le nom de variables, puis cliquez sur OK : les données s'affichent en mode tableur. Cliquez sur Données puis sur Définir des dates. Une boîte de dialogue Définir des dates s'ouvre (voir figure 2.6, page suivante).

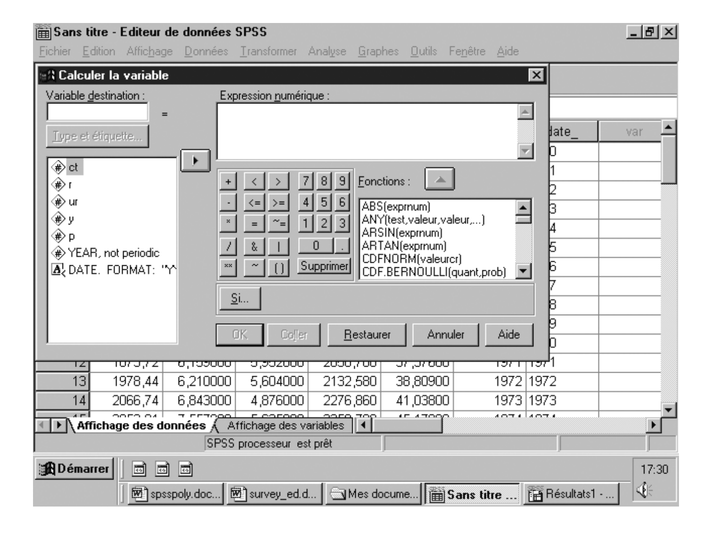
Sélectionnez Années dans Les observations sont, et saisissez 1960 dans Date de départ. Cliquez sur OK. La fenêtre Résultats1 – Viewer SPSS s'ouvre et signale que deux nouvelles variables ont été créées : YEAR_ et DATE_. Ces deux variables sont ici identiques et contiennent la liste des années successives. Réduisez la fenêtre : les données s'affichent

Figure 2.6



en mode tableur, avec deux colonnes supplémentaires, à savoir year_ et date_. Cliquez sur Transformer puis sur Calculer. La fenêtre Calculer la variable s'affiche comme à la figure 2.7.

Figure 2.7



Déplacez-vous pour rechercher LN(exprnum) dans la zone Fonctions. Double-cliquez sur LN(exprnum), ou bien sélectionnez LN(exprnum) et cliquez sur la flèche orientée vers le haut (voir figure 2.8, page ci-contre).

Sélectionnez ct, cliquez sur la flèche orientée vers la droite (voir figure 2.9, page ci-contre).

Saisissez LCT dans la zone Variable destination (voir figure 2.10, page ci-contre).

Cliquez sur OK : les données s'affichent en mode tableur avec une colonne supplémentaire, en l'occurrence lct. Répétez deux fois ces opérations pour créer successivement le logarithme de y et le logarithme de p. Les données apparaissent toujours en mode tableur, avec les colonnes supplémentaires ly et lp. Cliquez sur Transformer puis sur Créer la série chronologique. La fenêtre Créer la série chronologique s'ouvre. Double-cliquez sur lct, ou bien sélectionnez lct et cliquez sur la flèche orientée vers la droite. Sélectionnez Différence dans le menu déroulant Fonction, puis saisissez DLC dans la zone Nom et cliquez sur Remplacer (voir figure 2.11, page 64).



Figure 2.8

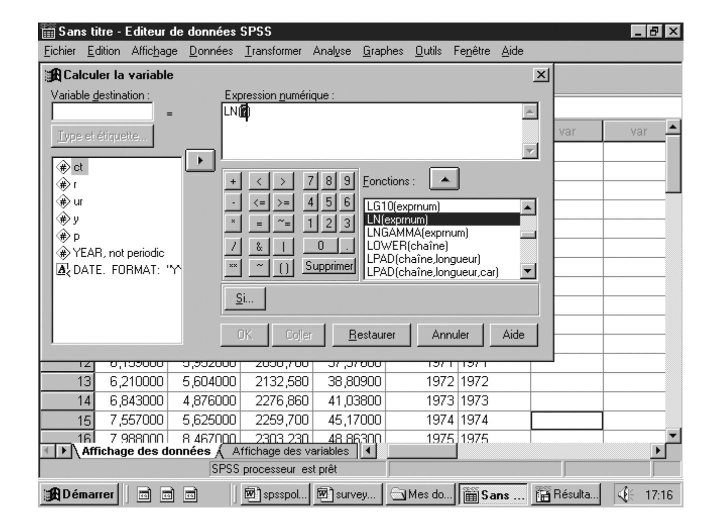


Figure 2.9

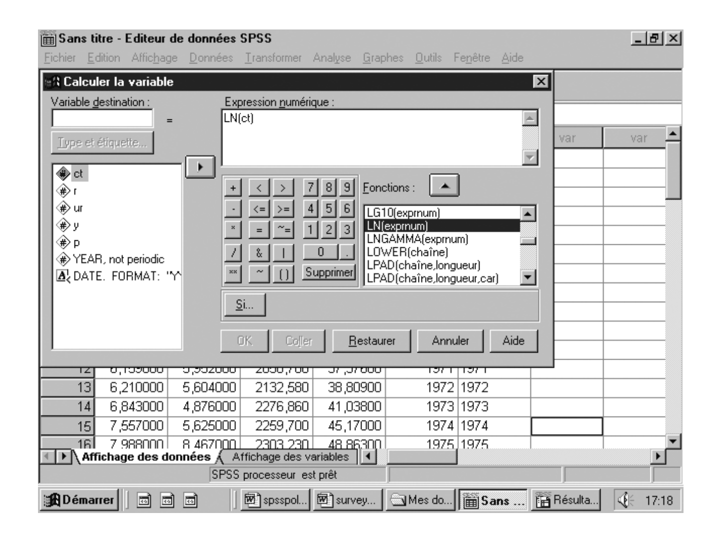
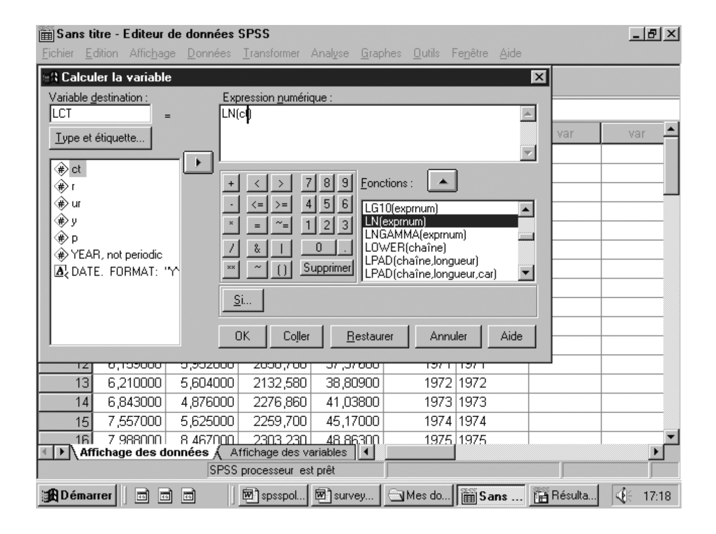


Figure 2.10



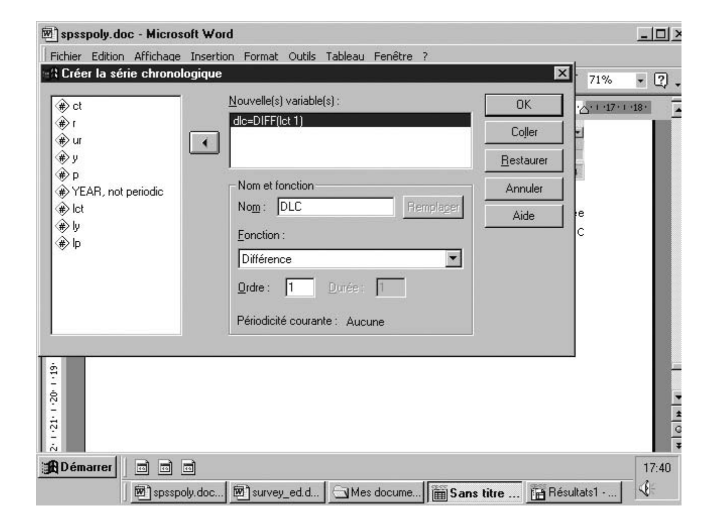


Figure 2.11

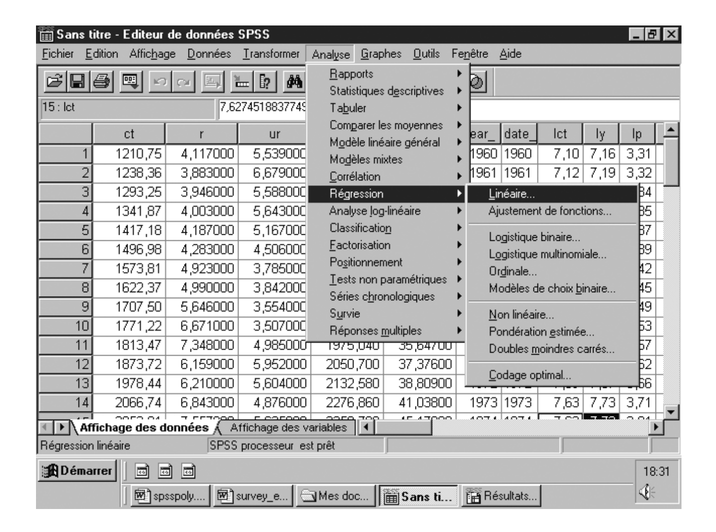


Figure 2.12

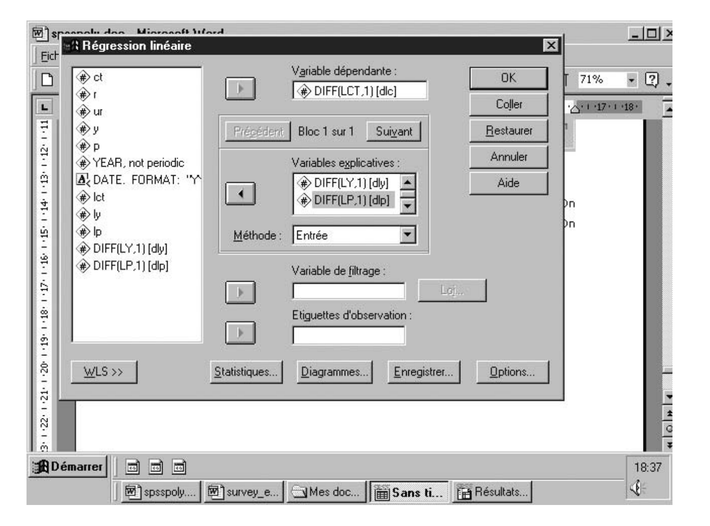


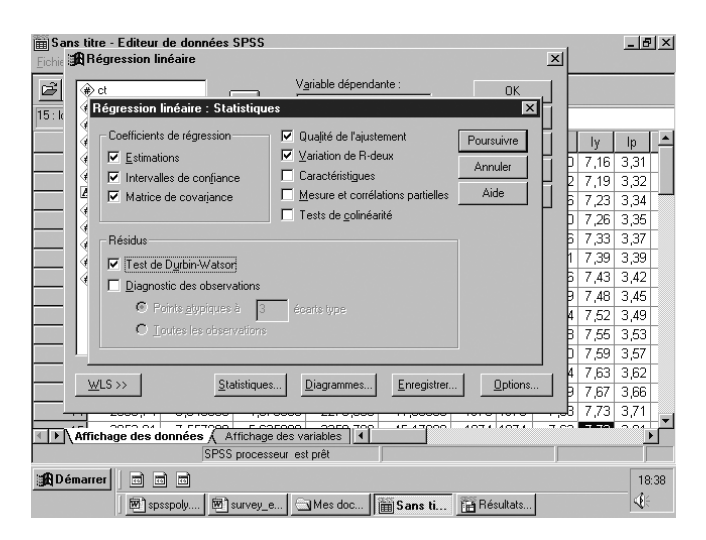
Figure 2.13

Cliquez sur OK : les données s'affichent en mode tableur avec une colonne supplémentaire, en l'occurrence dlc. Répétez les opérations pour calculer DLY et DLP. Les données s'affichent en mode tableur avec les colonnes supplémentaires dly et dlp. Cliquez sur Analyse dans la barre de menus horizontale puis sur Régression et enfin sur Linéaire (voir figure 2.12, page ci-contre).

Cliquez sur dlc puis sur la flèche vers la droite au niveau de la zone Variable dépendante. Cliquez sur dly puis sur la flèche vers la droite au niveau de la zone Variables explicatives. Cliquez sur dlp puis sur la flèche vers la droite, toujours au niveau de la zone Variables explicatives (voir figure 2.13, page ci-contre).

Sélectionnez Entrée dans la fenêtre Méthode. Cliquez sur Statistiques (voir figure 2.14).

Figure 2.14



Cliquez sur Poursuivre. Vous obtenez les résultats des figures 2.15 à 2.19, page suivante.

Figure 2.15

Variables introduites/éliminées[b]

Modèle	Variables introduites	Variables éliminées	Méthode
1	DIFF(LP,1),[a] DIFF(LY,1),[b]		Introduire

[a] Toutes variables requises introduites

[b] Variable dépendante : DIFF(LCT,1)

Figure 2.16

Récapitulatif du modèle[b]

Modèle	R	R^2	R^2 ajusté	Erreur standard de l'estimation	Changement dans les statistiques					Durbin Watson
					Variation de R^2	Variation de F	ddl 1	ddl 2	Modification de F signification	
1	0,822a	0,676	0,655	0,00975	0,676	32,292	2	31	0,000	1,953

[a] Valeurs prédites : (constantes), DIFF(LP,1), DIFF(LY,1)

[b] Variable dépendante : DIFF(LCT,1)

Exercices

Figure 2.17

ANOVA[a]

Modèle		Somme des carrés	ddl	Carré moyen	F	Signification
1	Régression	0,006	2	0,003	32,292	0,000[b]
	Résidu	0,003	31	0,000		
	Total	0,009	33			

a. Variable dépendante : DIFF(LCT,1)

b. Valeurs prédites : (constantes), DIFF(LP,1), DIFF(LY,1)

Figure 2.18

Coefficients[a]

Modèle	Coefficients non standardisés		Coefficients standardisés	t	Signification	Intervalle de confiance à 95 de B	
	B	Erreur standard	Bêta			Borne inférieure	Borne supérieure
1 (constante)	1,628E-02	0,006	0,006	2,534	0,017	0,003	0,029
DIFF(LY,1)	0,696	0,112	0,715	6,206	0,000	0,467	0,925
DIFF(LP,1)	-0,137	0,082	-0,192	-1,668	0,105	-0,304	0,030

a. Variable dépendante : DIFF(LCT,1)

Figure 2.19

Coefficients de corrélation[a]

Modèle			DIFF(LP,1)	DIFF(LY,1)
1	Corrélations	DIFF(LP,1)	1,000	0,461
		DIFF(LY,1)	0,461	1,000
	Covariances	DIFF(LP,1)	6,726E-03	4,243E-03
		DIFF(LY,1)	4,243E-03	1,258E-02

a. Variable dépendante : DIFF(LCT,1)

EXERCICE 4 RÉGRESSION LINÉAIRE AVEC EASYREG

Énoncé

Refaites l'exercice 1 en utilisant le logiciel Easyreg.

Solution

Généralités sur Easyreg

EasyReg est un logiciel économétrique avancé, téléchargeable sur le site **http://econ.la.psu.edu/~hbierens/EASYREG.HTM**. Il a été créé par le professeur Herman Bierens [(1)].

Démarrez Easyreg en cliquant sur son icône. Cliquez sur File, puis sur Choose an input file. Vous avez le choix entre :

- lire un fichier de données généré sous Excel et enregistré comme type de fichier CSV ;

1. H.J. Bierens, EasyReg International, Department of Economics, Pennsylvania State University, University Park, PA, 2002.

- lire un fichier enregistré dans un autre format ;
- continuer une session précédente, avec les données déjà utilisées au cours de cette session.

Pour les deux premières commandes, vous devez indiquer dans quel dossier de l'ordinateur démarrer la session (la procédure est expliquée plus loin).

Avant d'aller plus loin, sachez que lorsque vous démarrez Easyreg dans un dossier (répertoire ou folder) du disque dur, celui-ci crée dans ce dossier un sous-dossier (sous-répertoire ou sub-folder) EASYREG.DAT, avec d'une part des fichiers INPUT, INPUT1.RAN and INPUT2.RAN, et d'autre part plusieurs fichiers OUTPUT : un fichier texte avec les résultats appelé OUTPUT.TXT, et d'éventuels fichiers graphiques *.BMP. Si vous démarrez Easyreg à partir d'un répertoire contenant déjà un sous-répertoire EASYREG.DAT issu d'une session de travail précédente, le logiciel permet de continuer la session précédente. À la première utilisation de Easyreg, le répertoire de départ est C :\ ; ensuite, il démarre par défaut à partir du répertoire contenant les résultats de la session de travail la plus récente. Si vous travaillez simultanément sur plusieurs projets, il est utile de créer des (sous-)répertoires différents, et de démarrer systématiquement Easyreg dans le répertoire souhaité.

Une première manière d'organiser les données consiste à les placer dans un fichier CSV et à les délimiter par des points-virgules dans les versions francophones d'Excel où le symbole décimal utilisé est la virgule, ou par des virgules dans les versions où le symbole décimal utilisé est le point.

Si vous disposez de *k* variables ayant chacune n observations, organisez le fichier de données sous Excel de la manière suivante, où $x(i, j)$ est l'observation de date *i* pour la variable *j* :

```
Name(1)   Name(2) ...  Name(k)
x(1,1)    x(1,2)  ...  x(1,k)
x(2,1)    x(2,2)  ...  x(2,k)
.............................
x(n,1)    x(n,2)  ...  x(n,k)
```

Ensuite, enregistrez ce fichier au format CSV séparateur point-virgule. Puis ouvrez-le sous WordPad pour l'afficher :

```
Name(1);Name(2);...;Name(k)
x(1,1);x(1,2);...;x(1,k)
x(2,1);x(2,2);...;x(2,k)
.............................
x(n,1);x(n,2);...;x(n,k)
```

Les données $x(i, j)$ utilisent la virgule pour les décimales (format par défaut). Si vous optez pour ce format, le logiciel vous demandera ensuite quel est la balise à insérer pour signaler d'éventuelles données manquantes.

Le deuxième format de données possible est le suivant. Si vous disposez de *k* variables ayant chacune n observations et si vous souhaitez que la balise « *m* » représente d'éventuelles données manquantes, organisez le fichier de données de la manière suivante, où $x(i, j)$ est l'observation de date *i* pour la variable *j* :

```
k m
Name of variable 1
Name of variable 2
..................
Name of variable k
x(1,1) x(1,2) ... x(1,k)
x(2,1) x(2,2) ... x(2,k)
........................
x(n,1) x(n,2) ... x(n,k)
```

La balise « *m* » peut correspondre par exemple à −99999. Indiquez 0 si aucune donnée n'est manquante. Lorsqu'il s'agit de séries temporelles, les données ne peuvent manquer qu'aux observations extrêmes des variables. Pour des données transversales, elles peuvent manquer n'importe où.

Vous pouvez aussi réaliser ces fichiers directement sous WordPad, sans passer par Excel.

Lecture des données, transformation et estimation

Dans cet exemple, vous devez transformer le fichier Excel déjà utilisé, USA.XLS, en un fichier USA.TXT selon la méthode décrite précédemment.

Ouvrez le fichier USA.xls avec Excel et supprimez la première colonne de dates. Cliquez sur Fichier puis sur Enregistrer sous, et sélectionnez CSV (séparateur point-virgule) pour le type de fichier. Laissez Disque local C :\ dans la zone Enregistrer dans et conservez USA comme nom du fichier. Cliquez sur Enregistrer. Le logiciel fabrique alors un fichier USA.csv dans C :\.

Ouvrez Easyreg et cliquez successivement sur File/Choose an input file/Choose an Excel file in CSV format. Dans la fenêtre Guided tour qui s'ouvre, cliquez sur la réponse de votre choix. Si vous demandez de voir le Guided tour, vous devez ensuite réduire la fenêtre Guided tour on importing Excel files in CSV format pour pouvoir continuer à travailler avec Easyreg (et voir la fenêtre About CSV files). Dans la fenêtre About CSV files, cliquez sur Continue. Vous accédez à une fenêtre Find a CSV file. Sélectionnez Manually dans la zone Search options, C : dans Choose drive, et C :\ dans Choose drive/directory. Cochez Double click to choose file and continue. Double-cliquez sur C :\usa.csv dans Files found.

La fenêtre Data type of the CSV file s'affiche. Cochez sur Annual time series. Inscrivez 1960 dans Year of first observation et cliquez sur OK. Cliquez sur Confirm, puis sur Continue. La fenêtre Analyzing the CSV file s'affiche. Cliquez sur Code OK. La fenêtre Data written to files s'affiche. Cliquez sur Continue. La fenêtre Check the data s'ouvre. Cliquez sur Continue.

La fenêtre About the start-up folder EASYREG.DAT s'affiche. Cliquez sur Continue. La fenêtre Choose drive and folder to start-up EasyReg s'affiche. Double-cliquez sur C :\. Ensuite indiquez USA dans Make a new sub-folder in the Current folder, Enter the folder name et cliquez sur OK. Cliquez sur Start Easyreg in current folder puis sur OK.

Le logiciel revient au menu de départ. Cliquez successivement sur Menu/Input/Transform Variables pour calculer les différences logarithmiques des variables CT, Y et P, puisque le modèle à estimer est spécifié en fonction de ces transformations. Le logiciel affiche alors une liste de transformations possibles. Choisissez d'abord Logarithmic Transformations. Le logiciel affiche la liste des variables. Cliquer deux fois sur les variables à transformer : CT, Y et P. Une étoile se place devant chacune d'elles. Cliquez sur Selection OK. Le logiciel affiche la liste des nouvelles variables. Cliquez alors sur OK.

Vous revenez au menu des transformations. Cliquez alors sur Time series transformations. Une liste de transformations s'affiche. Cliquez sur Difference. Le logiciel propose un retard de une période pour la différence, donc $m = 1$. Cliquez une nouvelle fois sur Différence. Vous obtenez Select variables to be transformed avec la liste des variables existantes. Double-cliquez sur lN(CT), lN(YT) et LN(P) et cliquez sur Selection OK. Vous obtenez la liste des nouvelles variables en différence. Cliquez sur OK. Vous revenez au menu des transformations. Cliquez alors sur Done. Vous revenez au menu principal.

Pour estimer le modèle, cliquez successivement sur Menu/Single equation models/Linear regression models. Cliquez sur Not now. Le message Select all variables in your linear regression model s'affiche. Il faut donc sélectionner les noms des variables dépendante et explicatives, que celles-ci apparaissent dans le modèle avec un retard ou sans. Double-cliquez sur DIF1(LN(CT)), DIF1(LN(Y)) et DIF1(LN(P)). Cliquez sur Selection OK. Le logiciel vous demande si vous voulez utiliser un sous-ensemble d'observations. Cliquez sur No, puis sur Continue. Le message Select the Y variable in your OLS model s'affiche, avec la liste des variables. Sélectionnez la variable dépendante : double-cliquez sur DIF1(LN(CT)). Une étoile s'inscrit devant cette variable. Cliquez sur Continue.

Le logiciel affiche le nom de la variable choisie. Cliquez encore sur Continue. le message Select the independent X variables s'inscrit. Confirmez la sélection des variables explicatives. Le programme propose toutes les variables restantes, en l'occurrence DIF1(LN(Y)) et DIF1(LN(P)), en les faisant précéder d'une étoile. Si vous êtes d'accord avec cette sélection, cliquez sur Selection OK. Le logiciel vous demande d'indiquer si, parmi les variables explicatives, il faut ajouter un ou plusieurs retards de la variable dépendante. Ici, pour le modèle à estimer, ce n'est pas le cas. Ne sélectionnez donc rien et cliquez sur Skip all. Le logiciel vous demande s'il faut mettre dans le modèle la valeur contemporaine ou/et un ou plusieurs retards de la première variable explicative DIF1(LN(Y)). Ici, choisissez uniquement la valeur contemporaine : double-cliquez sur DIF1(LN(Y)) et cliquez sur Selection OK.

Le logiciel demande s'il faut mettre dans le modèle la valeur contemporaine ou/et un ou plusieurs retards de la deuxième variable explicative DIF1(LN(P)). De nouveau, choisissez uniquement la valeur contemporaine dans ce cas de figure : double-cliquez sur DIF1(LN(P)) puis cliquez sur Selection OK.

Le message Select deterministic X variables s'affiche. Cliquez sur No intercept si vous ne voulez pas de constante, et sur Time Trend pour ajouter un trend dans les variables explicatives. Ce n'est pas le cas ici. Cliquez donc sur Continue. Le logiciel demande s'il faut ajouter un polynôme de Tchebishev au modèle. Répondez par la négative dans cet exemple. Laissez donc $q = 0$ en cliquant sur q OK. Cliquez sur Continue à deux reprises. Le logiciel affiche les résultats :

```
Model variables :
y = DIF1[LN[CT]]
x(1) = DIF1[LN[Y]]
x(2) = DIF1[LN[P]]
x(3) = 1

Available observations : t = 2(=1961) -> 35(=1994)
= Chosen

OLS estimation results for Y = DIF1[LN[CT]]
Variables          OLS estimate        t-value  H.C.   t-value(*)
                                            s.e.      H.C. s.e.(*)
```

Exercices

```
                                            [p-value] [H.C. p-value](*)
x(1) = DIF1[LN[Y]]      .696018           6.206          6.395
                                        1.1216E-01    1.0884E-01
                                         [0.00000]     [0.00000]
x(2) = DIF1[LN[P]]     -.136784          -1.668         -1.910
                                        8.2010E-02    7.1610E-02
                                         [0.10541]     [0.05612]
x(3) = 1                .016282           2.534          2.537
                                        6.4249E-03    6.4187E-03
                                         [0.01654]     [0.01119]
(*) Based on White's heteroskedasticity consistent variance matrix.
The p-values are two-sided. The standard p-value = P(|t(31)|>|t-value|),
where t(31) is t distributed with 31 degrees of freedom.
The H.C. p-values are based on the normal approximation:
H.C. p-value = P(|u|>|H.C. t-value|), where u is N(0,1) distributed.

Standard error of the residuals:        97.549362E-004
Residual sum of squares (RSS):          29.499221E-004
Total sum of squares (TSS):             90.956092E-004
Overall F test: F(2,31):                         32.29
p-value = 0.00000
Significance levels:        10%        5%
Critical values:           2.48        3.3
Conclusions:             reject      reject
R-square:                                     0.675676
Adjusted R-square:                            0.654752
Effective sample size (n):                          34

Test for first-order autocorrelation:

Durbin-Watson test:                    1.953474

REMARK : A better way of testing for serial correlation
is to specify ARMA errors and then test the null
hypothesis that the ARMA parameters are zero.

Jarque-Bera/Salmon-Kiefer test:         .733446

Null hypothesis: The errors are normally distributed
Null distribution: Chi-square(2))
         p-value = 0.69300
         Significance levels:              10%        5%
         Critical values:                 4.61      5.99
         Conclusions:                   accept    accept

Breusch-Pagan test:                     .179704

Null hypothesis : The errors are homoskedastic
Null distribution : Chi-square(2)
         p-value = 0.91407
         Significance levels:              10%        5%
         Critical values:                 4.61      5.99
         Conclusions:                   accept    accept
```



```
If the model is correctly specified, in the sense that the conditional
expectation of the model error u relative to the X variables and all
lagged dependent (y) variables and lagged X variables equals zero,
then the OLS parameter estimators b(1),...,b(3), minus their true values,
are (asymptotically) jointly normally distributed with zero mean vector
and variance matrix:
 1.25790E-002   4.24264E-003 -5.87975E-004
 4.24264E-003   6.72559E-003 -4.39615E-004
 -5.87975E-004 -4.39615E-004  4.12796E-005
provided that the conditional variance of the model error u is constant
(u is homoskedastic), or
 1.18457E-002   5.07015E-003 -6.19792E-004
 5.07015E-003   5.12796E-003 -4.01183E-004
 -6.19792E-004 -4.01183E-004  4.11998E-005
if the conditional variance of the model error u is not constant
(u is heteroskedastic).
```

Le logiciel affiche donc les coefficients estimés (OLS estimate), suivis de leur statistique *t* (*t* value), de leur écart type estimé (s.e.), de la probabilité critique du test de l'hypothèse que le coefficient vrai correspondant est nul (*p* value). À la colonne suivante le logiciel affiche les écarts types calculés selon la méthode de White (voir chapitre 3) robuste à une hétéroscédasticité éventuelle des erreurs (H.C. s.e.) ainsi que les statistiques *t* (H.C. *t* value) et les probabilités critiques correspondantes (H.C. *p* value). Ces valeurs modifiées sont quasiment identiques aux valeurs calculées normalement, ce qui démontre l'absence de problème d'hétéroscédasticité des erreurs. L'avantage d'Easyreg est qu'il rappelle pour chaque test l'hypothèse précise qui est testée (null hypothesis) et qu'il donne explicitement la conclusion de l'interprétation.

Références bibliographiques

[ANS 1961] F.J. Anscombe, Examination of Residuals, dans *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 4, University of California Press, Berkeley, p. 1–36, 1961.

[BER 1981] A.K. Bera, C.M. Jarque, *An Efficient Large-Sample Test for Normality of Observations and Regression Residuals*, Australian National University Working Papers in Econometrics, 40, Canberra, 1981.

[BOX 1970] G.E.P. Box, D.A. Pierce, Distribution of Residual Autocorrelations in Autoregressive-Integrated-Moving Average Time Series Models, dans *Journal of American Statistical Association*, 65, p. 1509–1526, 1970.

[BRE 1981] T.S. Breusch, L.G. Godfrey, A Review of Recent Work on Testing for Autocorrelation in Dynamic Simultaneous Models, dans *Macroeconomic Analysis : Essays in Macroeconomics and Econometrics*, Edition D. Currie, R. Nobay, et D. Peel, Croom Helm, London, 1981.

[BRE 1980] T.S. Breusch, A.R. Pagan, The Lagrange Multiplier Test and its Application to Model Specifications in Econometrics, dans *Review of Economic Studies*, 47, p. 239–253, 1980.

[BRE 1979] T.S. Breusch, A.R. Pagan, A Simple Test for Heteroscedasticity and Random Coefficient, dans *Econometrica*, 47, p. 1287–1294, 1979.

[BRO 1975] R.L. Brown, J. Durbin, J.M. Evans, Techniques for Testing the Constancy of Regression Relations Over Time (with discussion), dans *Journal of the Royal Statistical Society B*, 37, p. 149–192, 1975.

[CHO 1960] G.C. Chow, Test of Equality Between Sets of Coefficients in Two Linear Regressions, dans *Econometrica*, 28, p. 591–605, 1960.

[DAV 1993] R. Davidson, J.G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, Oxford, 1993.

[DUR 1971] Durbin, J., 1971, Boundary-Crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test, Journal of Applied Probability 8, p. 431–453.

[DUR 1970] J. Durbin, Testing for Serial Correlation in Least-Squares Regression when Some of the Regressors are Lagged Dependent Variables, dans *Econometrica*, 38, p. 410–421, 1970.

[DUR 1969] Durbin, J., 1969, Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least Squares Residuals, Biometrika.

[DUR 1951] J. Durbin, G. S. Watson, Testing for Serial Correlation dans Least Squares Regression II, dans *Biometrika*, 38, p. 159–178, 1951.

[DUR 1950] J. Durbin, G.S. Watson, Testing for Serial Correlation dans Least Squares Regression I, dans *Biometrika*, 37, p. 409–428, 1950.

[ENG 1982] R.F. Engle, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, dans *Econometrica*, 50, p. 987–1007, 1982.

[GOD(a) 1978] L.G. Godfrey, Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables, dans *Econometrica*, 46, p. 1293–1301, 1978.



[GOD(b) 1978] L.G. Godfrey, Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables, dans *Econometrica*, 46, p. 1303–1310, 1978.

[GRE 1993] W. H. Greene, *Econometric Analysis*, deuxième edition, Macmillan, New York, 1993.

[HEN 1995] D. Hendry, *Dynamic Econometrics*, Oxford University Press, 1995.

[JAR 1987] C.M. Jarque, A.K. Bera, A Test for Normality of Observations and Regression Residuals, dans *International Statistical Review*, 55, p. 163–172, 1987.

[JAR 1980] C.M. Jarque, A.K. Bera, Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals, dans *Economics Letters*, 6, p. 255–259, 1980.

[JOH 1997] J. Johnston, J. DiNardo, *Econometric Methods*, McGraw Hill, 1997.

[LJU 1978] G.M. Ljung, G.E.P. Box, On a Measure of Lack of Fit dans Time Series Models, dans *Biometrika*, 65, p. 297–303, 1978.

[RAM 1970] J.B. Ramsey, Models, Specification Error and Inference : A Discussion of Some Problems in Econometric Methodology, dans *Bulletin of the Oxford Institute of Economics and Statistics*, 32, p. 301–318, 1970.

[RAM 1969] J.B. Ramsey, Test for Specification Errors in Classical Linear Least Squares Regression Analysis, dans *Journal of the Royal Statistical Society B*, p. 350–371, 1969.

[RAY 1994] R.K. Rayner, The small-sample power of Durbin's h test revisited, dans *Computational Statistics and Data Analysis*, 17, January, p. 87–94, 1994.

[SPA 1986] A. Spanos, *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge, 1986.

[WHI 1980] H. White, A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, dans *Econometrica*, 48, p. 817–838, 1980.

Chapitre 3

Compléments sur les modèles linéaires

Compléments sur les modèles linéaires

Ce chapitre aborde plusieurs points relatifs à l'inférence statistique basée sur un modèle linéaire. Il traite des tests d'hypothèses sur les coefficients du modèle, de l'estimation du modèle linéaire par maximum de vraisemblance, de la multicolinéarité, de l'utilisation de variables indicatrices, des estimateurs de variance robustes, de l'estimateur des moindres carrés généralisés, des modèles linéaires sur des données en coupe instantanée et de la méthode des variables instrumentales.

1 Tests d'hypothèses linéaires sur plusieurs coefficients

1.1 Test général de restrictions linéaires sur les coefficients

On suppose que Y et X sont liés par un modèle linéaire $Y = X\beta + u$, où u est indépendant de X et est un bruit blanc normal : $u \sim N(0, \Sigma_u)$ où $\Sigma_u = \sigma_u^2 I_n$. On souhaite tester un ensemble de restrictions linéaires portant chacune sur un ou plusieurs coefficients du vecteur β.

Exemple

On veut tester la véracité simultanée des deux restrictions suivantes : $\beta_2 = 1$ et $\beta_3 = \beta_4$ dans le modèle :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$$

Pour cela, trois étapes :

1. On estime le modèle par MCO sans aucune restriction, et l'on garde la somme des carrés des résidus ainsi obtenue, notée SE.
2. On estime le modèle par MCO *en imposant les restrictions*, et l'on garde la somme des carrés des résidus ainsi obtenue, notée $SE2$.
3. On montre que, *si les restrictions sont vraies*, alors :

$$\frac{(n-k)\,(SE2 - SE)}{mSE} \sim F(m, n-k) \tag{3.1}$$

et il suffit de calculer la valeur de ce test et de la comparer aux valeurs critiques d'une table de la distribution de Ficher à m degrés de liberté au numérateur et $n-k$ degrés de liberté au dénominateur. Il s'agit d'un test de petit échantillon, donc d'un test exact quelle que soit la taille de l'échantillon, si le terme d'erreur est normal. C'est en fait une transformation monotone du test du rapport de vraisemblance des mêmes restrictions, mais ici la distribution est exacte puisqu'on a fait l'hypothèse que le terme d'erreur est normal.

Exemple (suite)

Pour tester les deux restrictions de l'exemple précédent, on estime donc le modèle sans restrictions :

$$Y_t = \hat{\beta}_1^{MCO} + \hat{\beta}_2^{MCO} X_{2t} + \cdots + \hat{\beta}_4^{MCO} X_{4t} + e_t$$

et l'on garde la valeur $SE = e'e = \sum_{t=1}^{n} e_t^2$. Puis on estime le modèle avec restrictions :

$$Y_t - X_{2t} = \hat{\beta}_1^{MCO} + \hat{\beta}_3^{MCO}(X_{3t} + X_{4t}) + \hat{e}_t$$

et l'on garde la valeur $SE2 = \hat{e}'\hat{e} = \sum_{t=1}^{n} \hat{e}_t^2$. Puis on calcule :

$$\frac{(n-k)\,(SE2 - SE)}{mSE}$$

et l'on compare la valeur obtenue aux valeurs critiques d'une table $F(2, n-4)$, ou bien on utilise un logiciel pour obtenir la probabilité critique.

1.2 TEST D'UNE OU DE PLUSIEURS RESTRICTIONS LINÉAIRES SUR LES COEFFICIENTS

On suppose que Y et X sont liés par un modèle linéaire $Y = X\beta + u$, où u est indépendant de X, est un bruit blanc : $\Sigma_u = \sigma_u^2 I_n$ et a une distribution normale. Un ensemble de m restrictions linéaires sur les coefficients β peut être représenté de manière générale par :

$$R\beta = r \tag{3.2}$$

où R, β et r sont des matrices de type $m \times k$ pour R, $k \times 1$ pour β, et $m \times 1$ pour r.

Exemple (suite)

Les restrictions $\beta_2 = 1$ et $\beta_3 = \beta_4$ peuvent être représentées par :

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Sous l'hypothèse H_0 que ces restrictions sont vraies, c'est-à-dire que $R\beta = r$, on montre aisément que :

$$\frac{W}{m} = (R\beta - r)' \left(R\left(X'X\right)^{-1} R'\right)^{-1} (R\beta - r)\, \frac{1}{m\left(\dfrac{e'e}{n-k}\right)} \sim F\,(m, n-k) \tag{3.3a}$$

Il s'agit d'un test de Wald *modifié* par une division par m, qui nécessite uniquement une estimation du modèle sans restrictions. On montre que la valeur obtenue est identique à celle du test calculé selon la formule de la section précédente pour les mêmes restrictions, et qui nécessite les estimations du modèle sans restrictions et avec restrictions.

Dès lors que le terme d'erreur est un bruit blanc et normal, il s'agit d'un test de petit échantillon, donc d'un test exact quelle que soit la taille de l'échantillon.

Si u n'a pas une distribution normale, le test est quand même asymptotique. Puisque $\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{L} N(0, \sigma_u^2 Q^{-1})$ et $p\lim \dfrac{X'X}{n} = Q$, on montre aisément que :

$$W = (R\beta - r)' \left(R\left(X'X\right)^{-1} R'\right)^{-1} (R\beta - r)\, \frac{1}{\left(\dfrac{e'e}{n-k}\right)} \xrightarrow{L} \chi^2_m \tag{3.3b}$$

Il s'agit d'un vrai test de Wald. Bien que la distribution asymptotique de W soit χ^2, la distribution F appliquée à $\dfrac{W}{m}$ fournit en pratique une meilleure approximation pour n petit. Pour tester l'hypothèse $R\beta = r$, il suffit donc de calculer le membre de gauche de l'expression précédente, de diviser le résultat par m, et de comparer le tout aux valeurs critiques d'une table de la distribution Fischer à m degrés de liberté au numérateur et $n-k$ au dénominateur. Pour réaliser ce test, on n'a besoin uniquement d'estimer le modèle non contraint par MCO, et ensuite de calculer la valeur du test par la formule précédente.

1.3 TEST D'UNE SEULE RESTRICTION LINÉAIRE SUR LES COEFFICIENTS

On souhaite tester une seule hypothèse H_0 de restriction linéaire sur les coefficients, qui peut se formaliser par :

$$w'\beta = x \tag{3.4}$$

où w est un vecteur $k \times 1$ et x un scalaire 1×1.

Exemple

La restriction $\beta_3 = \beta_4$ peut être représentée par $(0\ 0\ -1\ 1)\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\end{pmatrix} = (0)$. Dans ce cas, $r = 0$ et

$w = \begin{pmatrix}0\\ 0\\ -1\\ 1\end{pmatrix}$.

On montre que sous H_0, donc si $w'\beta = x$ est vraie, alors :

$$\frac{w'\hat{\beta}^{MCO} - x}{\sqrt{\dfrac{e'e}{n-k}}\sqrt{w'\,(X'X)^{-1}\,w}} \sim t_{n-k} \tag{3.5}$$

Pour tester l'hypothèse $w'\beta = x$, il suffit d'estimer le modèle sans restrictions, $Y = X\beta + u$, par MCO. On obtient le modèle estimé $Y = X\beta^{MCO} + e$. On calcule alors la valeur de la formule du test et on la compare aux valeurs critiques d'une distribution t_{n-k}. Il s'agit là d'un cas particulier du test précédent : celui où $m = 1$. On sait en effet qu'une variable aléatoire ayant une distribution $F(1, n-k)$ a une racine carrée qui a une distribution t_{n-k}.

1.4 TESTS DE VALEURS PARTICULIÈRES POUR TOUS LES COEFFICIENTS

On souhaite tester l'hypothèse que le vecteur β est égal à un vecteur de valeurs particulières $\bar{\beta}$. On sait que, pour les valeurs « vraies inconnues » β de la réalité, $\dfrac{\left(\hat{\beta}^{MCO} - \beta\right)' X'X \left(\hat{\beta}^{MCO} - \beta\right)(n-k)}{(e'e)\,k} \sim F(k, n-k)$ (voir chapitre 2). Cela implique que si $\beta = \bar{\beta}$, alors $\dfrac{\left(\hat{\beta}^{MCO} - \bar{\beta}\right)' X'X \left(\hat{\beta}^{MCO} - \bar{\beta}\right)(n-k)}{(e'e)\,k} \sim F(k, n-k)$. Pour vérifier l'hypothèse que $\bar{\beta} = \beta$, il suffit donc d'estimer le modèle linéaire $Y = X\beta + u$ sans restrictions. On obtient un modèle estimé $Y = X\beta^{MCO} + e$, on calcule la formule du test et on la compare aux valeurs critiques d'une distribution $F(k, n-k)$. Dès lors que le terme d'erreur est un bruit blanc et normal, il s'agit d'un test de petit échantillon, donc d'un test exact quelle que soit la taille de l'échantillon, c'est-à-dire même avec peu d'observations. C'est un cas particulier du test de Wald modifié, ou test F, de plusieurs restrictions linéaires sur les coefficients lorsque le terme d'erreur est un bruit blanc.



2 Estimation du modèle linéaire par maximum de vraisemblance

L'objet de cette section est d'expliquer comment on peut estimer un modèle linéaire du type $Y = X\beta + u$ par la technique du maximum de vraisemblance [1], sous l'hypothèse que u est indépendant de X et que $u \sim N(0, \Sigma_u)$, où $\Sigma_u = \sigma_u^2 I_n$ (u bruit blanc). Les hypothèses faites sur le terme d'erreur impliquent les égalités suivantes :

$$f_u(u) = f_{u_1 u_2 \ldots u_n}(u_1, u_2, \ldots u_n) = f_{u_1}(u_1) f_{u_2}(u_2) \ldots f_{u_n}(u_n) = \prod_{t=1}^{n} f_{u_t}(u_t) \qquad (3.6)$$

$$f_{u_t}(u_t) = \frac{1}{\sigma_u \sqrt{2\pi}} e^{-\frac{u_t^2}{2\sigma_u^2}} \quad \text{pour tout } t \qquad (3.7)$$

On observe la réalisation du vecteur Y et celle de la matrice X. La fonction de vraisemblance de l'échantillon est donc $f_{Y, X_2, \ldots X_n}(Y, X_2, \ldots X_n)$. Si l'on raisonne conditionnellement à X, la fonction de vraisemblance de l'échantillon est $f_{Y|X_2, \ldots X_n}(Y|X_2, \ldots X_n)$, notée simplement $f_{Y|X}(Y|X)$. Puisque Y est une fonction de u, la distribution de Y conditionnellement à X doit être calculée à partir de la distribution de u conditionnellement à X, qui est égale à la distribution marginale de u, puisque u et X sont indépendants : $f_u(u|X) = f_u(u)$. Il suffit donc d'appliquer la formule habituelle permettant d'obtenir la fonction de densité d'une variable aléatoire, qui est fonction d'une autre variable aléatoire dont on connaît la densité. Puisque le Jacobien vaut 1, on obtient :

$$f_{Y|X_2, \ldots X_n}(Y|X_2, \ldots X_n) = \prod_{t=1}^{n} \frac{1}{\sigma_u \sqrt{2\pi}} e^{-\frac{\left(Y_t - \beta_1 - \sum_{i=2}^{k} \beta_i X_{it}\right)^2}{2\sigma_u^2}} \underset{\text{déf}}{=} L\left(\beta, \sigma_u^2\right) \qquad (3.8)$$

Par conséquent, le logarithme de la fonction de vraisemblance est fourni par :

$$\begin{aligned} \ln L\left(\beta, \sigma_u^2\right) &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma_u^2 - \sum_{t=1}^{n} \frac{\left(Y_t - \beta_1 - \sum_{i=2}^{k} \beta_i X_{it}\right)^2}{2\sigma_u^2} \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln - \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma_u^2} \end{aligned} \qquad (3.9)$$

Les estimateurs de maximum de vraisemblance de β et σ_u^2 sont les solutions du système d'équations formé par les deux conditions de premier ordre :

$$\begin{aligned} 0 &= \frac{\partial \ln L\left(\beta, \sigma_u^2\right)}{\partial \beta} \\ 0 &= \frac{\partial \ln L\left(\beta, \sigma_u^2\right)}{\partial \sigma_u^2} \end{aligned} \qquad (3.10)$$

Ces solutions sont :

$$\begin{aligned} \hat{\beta}^{MV} &= \left(X'X\right)^{-1} X'Y \\ \hat{\sigma}_u^{2MV} &= \frac{e'e}{n} \end{aligned} \qquad (3.11)$$

1. Pour quelques rappels utiles sur l'estimation par la méthode du maximum de vraisemblance, le lecteur peut se référer par exemple au chapitre 5 du livre de Patrick Roger, Probabilités, statistique et processus stochastiques, publié chez Pearson Education France dans la même collection.

Les conditions de deuxième ordre pour un maximum sont respectées. L'estimateur de maximum de vraisemblance de β, noté $\hat{\beta}^{MV}$, est ainsi égal à l'estimateur des moindres carrés ordinaires de β, noté $\hat{\beta}^{MCO}$. Par contre, l'estimateur de maximum de vraisemblance de σ_u^2, noté $\hat{\sigma}_u^{2MV}$, n'est pas égal à l'estimateur des moindres carrés ordinaires de σ_u^2, noté $\hat{\sigma}_u^{2MCO}$, car $\hat{\sigma}_u^{2MV} = \frac{n-k}{n}\sigma_u^{2MCO}$. Bien sûr, cette différence est d'autant plus petite que n est grand, et elle devient insignifiante quand n est très grand.

La matrice d'information est :

$$R = -E\left(\frac{\partial \ln L\left(\beta, \sigma_u^2\right)}{\partial \begin{pmatrix} \beta \\ \sigma_u^2 \end{pmatrix} \partial \begin{pmatrix} \beta \\ \sigma_u^2 \end{pmatrix}}\right) = \begin{pmatrix} \sigma_u^{-2}\left(X'X\right) & 0 \\ 0 & \frac{n}{2\sigma_u^4} \end{pmatrix} \qquad (3.12)$$

Par conséquent, la borne de Rao Cramer est :

$$R^{-1}(\beta, \sigma_u^2) = \begin{pmatrix} \sigma_u^2\left(X'X\right)^{-1} & 0 \\ 0 & \frac{2\sigma_u^4}{n} \end{pmatrix} \qquad (3.13)$$

L'estimateur de MCO de β, égal à l'estimateur de maximum de vraisemblance de β, a une matrice de variance et de covariance égale à la borne de Rao Cramer. Il est donc forcément le plus précis de tous les estimateurs sans biais. Il est en outre efficient. On remarque que $\hat{\sigma}_u^{2MV}$ est biaisé : $E\left(\hat{\sigma}_u^{2MV}\right) \neq \sigma_u^2$ alors que $\hat{\sigma}_u^{2MCO}\sigma_u^2$ est sans biais : $E\left(\hat{\sigma}_u^{2MCO}\right) = \sigma_u^2$. Toutefois $\hat{\sigma}_u^{2MCO}$ a une variance supérieure à la borne de Rao Cramer : $V\left(\hat{\sigma}_u^{2MCO}\right) = \frac{2\sigma_u^4}{n-4} > \frac{2\sigma_u^4}{n}$. Cependant, pour un grand échantillon, on se rapproche de la borne : $\lim_{n\to\infty} V\left(\hat{\sigma}_u^{2MCO}\right) = \frac{2\sigma_u^4}{n}$. Par conséquent, $\hat{\sigma}_u^{2MCO}$ est asymptotiquement efficient.

Test du rapport de vraisemblance. La fonction de vraisemblance fournit un test asymptotique pratique et général pour tester simultanément plusieurs contraintes portant chacune sur un ou plusieurs coefficients du modèle linéaire. Le principe est d'estimer d'abord le modèle linéaire sans restrictions sur ses coefficients. La valeur de la fonction de vraisemblance maximisée de ce modèle est notée L_1. Ensuite, on estime le modèle linéaire en imposant à ses coefficients toutes les restrictions que l'on souhaite tester. La valeur de la fonction de vraisemblance maximisée de ce modèle contraint est notée L_0. On montre que, si les restrictions sont vraies, la valeur absolue de $2\left(\ln(L_1) - \ln(L_0)\right)$ suit asymptotiquement une loi Chi-2 à p degrés de liberté, où p est le nombre de restrictions testées sur les coefficients. On rejette les restrictions quand la valeur de ce test est supérieure aux valeurs critiques à un seuil de 5 % ou de 1 % d'une loi de Chi-2 à p degrés de liberté. En d'autres termes, c'est lorsque les restrictions provoquent une forte chute de la fonction de vraisemblance : L_1 est très supérieure à L_0, ce qui rend la différence de leurs logarithmes trop grande. Le test du rapport de vraisemblance est asymptotique, c'est-à-dire qu'il suit une distribution de Chi-2 quand le nombre d'observations tend vers l'infini. Pour des échantillons réduits, sa distribution n'est qu'approximative ; des tests de petit échantillon peuvent être plus fiables.



3 Multicolinéarité

On parle de *multicolinéarité* lorsqu'il y a des corrélations assez élevées entre certaines variables explicatives, ce qui affecte la matrice $(X'X)^{-1}$. On peut calculer l'estimateur des moindres carrés ordinaires, mais il est d'autant plus imprécis que la multicolinéarité est forte. Plus celle-ci augmente, plus les variances des coefficients estimés croissent aussi. Si des variables explicatives sont très liées linéairement, il est difficile de mesurer l'impact respectif de chacune d'elles sur la variable à expliquer. Souvent, en cas de corrélation forte entre deux variables explicatives, on ne peut rejeter l'hypothèse de nullité de leurs coefficients pris individuellement, alors que l'on rejette l'hypothèse de nullité conjointe. L'observation de ce phénomène indique la présence de multicolinéarité.

La multicolinéarité est dite *parfaite* quand une colonne de la matrice X est une combinaison linéaire d'une ou de plusieurs autres colonnes de X, en d'autres termes quand une variable explicative est une combinaison linéaire exacte d'autres variables explicatives. Dans ce cas, le rang de X est inférieur à k, ce qui implique que le rang de $X'X$ est également inférieur à k. La matrice $X'X$ n'est donc pas inversible ; il est impossible de calculer l'estimateur des moindres carrés ordinaires.

4 Variables indicatrices

Une variable est dite **indicatrice** (ou **dummy**) quand elle est artificielle ; on lui attribue des valeurs particulières aux différentes observations, pour faire varier la valeur de la constante de manière déterminée, en fonction des observations. En économétrie des séries temporelles, on emploie le plus souvent des variables indicatrices d'impulsion, de saut et de saison.

4.1 Variable indicatrice d'impulsion

Une variable indicatrice d'impulsion est telle que toutes ses observations valent 0, sauf une observation, qui vaut 1. Si t_i est la date de l'impulsion, alors :

$$\begin{aligned} DU_t &= 0 \quad \text{si} \ \ t \neq t_i \\ DU_t &= 1 \quad \text{si} \ \ t = t_i \end{aligned} \tag{3.14}$$

Lorsqu'on introduit une telle variable parmi les k variables explicatives d'un modèle linéaire, celui-ci devient :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-1} X_{kt-1} + \beta_k DU_t + u_t \quad \text{pour} \ \ t = 1 \ldots n \tag{3.15}$$

et :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-1} X_{kt-1} + u_t \quad \text{pour} \ \ t < t_i \ \ \text{et} \ \ t > t_i \tag{3.16}$$

et

$$Y_t = (\beta_1 + \beta_k) + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-1} X_{kt-1} + u_t \quad \text{pour} \ \ t = t_i \tag{3.17}$$

La constante du modèle vaut β_1 en toute période, sauf à la période t_i où elle vaut $\beta_1 + \beta_k$. On utilise une variable indicatrice d'impulsion lorsque la variable dépendante présente un saut à partir d'une date bien particulière, qui ne peut pas être expliqué par un saut comparable, à partir de cette date, d'une variable explicative.

4.2 VARIABLE INDICATRICE DE SAUT

Une variable indicatrice de saut est telle que toutes ses observations valent 0 avant une certaine date, et toutes ses observations valent 1 à partir de cette date :

$$DU_t = 0 \quad \text{si} \quad t < t_i \qquad (3.18)$$
$$DU_t = 1 \quad \text{si} \quad t \geqslant t_i$$

Lorsqu'on introduit une telle variable parmi les k variables explicatives d'un modèle linéaire, celui-ci devient :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-1} X_{kt-1} + \beta_k DU_t + u_t \quad \text{pour} \quad t = 1 \ldots n \quad (3.19)$$

ou encore :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-1} X_{(k-1)t} + u_t \quad \text{pour} \quad t < t_i \qquad (3.20)$$

et

$$Y_t = (\beta_1 + \beta_k) + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-1} X_{(k-1)t} + u_t \quad \text{pour} \quad t \geqslant t_i \quad (3.21)$$

La constante du modèle vaut β_1 en toute période avant la date t_i, et $\beta_1 + \beta_k$ à la date t_i et ultérieurement. On utilise une variable indicatrice d'impulsion lorsque la variable dépendante présente un saut à partir d'une date bien particulière, qui ne peut pas être expliqué par un saut comparable, à partir de cette date, d'une variable explicative.

4.3 VARIABLE INDICATRICE SAISONNIÈRE

En données mensuelles, on introduit des variables indicatrices saisonnières pour obtenir une constante ayant les propriétés suivantes :

- Elle est différente pour chacun des douze mois d'une même année.
- Elle est identique pour un même mois, quelle que soit l'année.

On définit des variables indicatrices $DU1, DU2 \ldots DU11$ de la manière suivante, quelle que soit l'année :

$DU1_t = 1$ si t est un mois de janvier.

$DU1_t = 0$ si t n'est pas un mois de janvier ou de décembre.

$DU1_t = -1$ si t est un mois de décembre.

$DU2_t = 1$ si t est un mois de février.

$DU2_t = 0$ si t n'est pas un mois de février ou de décembre.

$DU2_t = -1$ si t est un mois de décembre.

...

$DU11_t = 1$ si t est un mois de novembre.

$DU11_t = 0$ si t n'est pas un mois de novembre ou de décembre.

$DU11_t = -1$ si t est un mois de décembre.

Lorsqu'on introduit de telles variables parmi les k variables explicatives d'un modèle linéaire, celui-ci devient :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-11} X_{(k-11)t} + \beta_{k-10} DU1_t + \beta_{k-9} DU1_t + \cdots + \beta_k DU11_t + u_t \quad \text{pour} \quad t = 1 \ldots n \tag{3.22}$$

et :

$$Y_t = (\beta_1 + \beta_{k-10}) + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-1} X_{kt-1} + u_t$$
si t est un mois de janvier,
$$Y_t = (\beta_1 + \beta_{k-9}) + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-11} X_{kt-11} + u_t$$
si t est un mois février,

...

$$Y_t = \left(\beta_1 - (\beta_{k-10} + \beta_{k-9} + \cdots + \beta_k)\right) + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_{k-11} X_{(k-11)t} + u_t \quad \text{si } t \text{ est un mois de décembre.}$$

On utilise des variables indicatrices saisonnières lorsque la variable dépendante présente un cycle saisonnier qui ne peut être expliqué par l'évolution des variables explicatives, ces dernières n'ayant pas de cycle comparable. On évite ainsi que cette composante cyclique inexpliquée se retrouve dans les résidus calculés du modèle estimé, qui sera autocorrélé. La variation de la valeur de la constante, selon les mois de l'année, prend en compte le cycle saisonnier de la variable dépendante, palliant le fait que l'évolution des variables explicatives ne peut l'expliquer.

Remarque

Pour les données trimestrielles, le principe est le même, mais on définit seulement trois variables indicatrices.

5 Estimateurs de variance robustes

5.1 Estimateurs de variance robustes à l'hétéroscédasticité

Si le terme d'erreur u est hétéroscédastique, la vraie matrice $\Sigma_{\hat{\beta}^{MCO}|X}$ de variance et de covariance des estimateurs de MCO n'est pas égale à $\sigma_u^2 \left(X'X\right)^{-1}$ et ne peut donc être estimée par $\hat{\Sigma}_{\hat{\beta}^{MCO}|X} = \dfrac{e'e}{n-k}(X'X)^{-1}$, qui est la formule utilisée par défaut par les logiciels d'économétrie. Le plus fréquemment, la forme de l'hétéroscédasticité est inconnue, tout comme la formule précise de $\Sigma_{\hat{\beta}^{MCO}|X}$. Il est donc impossible d'établir un estimateur sur mesure de cette matrice. White [WHI 1980] a toutefois défini un estimateur convergent de $\Sigma_{\hat{\beta}^{MCO}|X}$ quelle que soit la forme de l'hétéroscédasticité. MacKinnon et White [MAC 1985] en ont proposé une version corrigée pour les degrés de liberté, mieux adaptée pour les petits échantillons [(1)]

$$\hat{\Sigma}^{White^{adj}}_{\hat{\beta}^{MCO}|X} = \frac{n}{n-k} \left(X'X\right)^{-1} \left(\sum_{t=1}^{n} e_t^2 x_t x_t'\right) \left(X'X\right)^{-1}$$

1. La formule originale de White [WHI 1980] est $\hat{\Sigma}^{White}_{\hat{\beta}^{MCO}|X} = (X'X)^{-1} \left(\sum_{t=1}^{n} e_t^2 x_t x_t'\right) (X'X)^{-1}$.

où x_t est un vecteur dont les éléments sont les observations de la ligne t de la matrice X. Les racines carrées des éléments de la diagonale principale de X sont donc des estimateurs convergents des écarts types (standard errors) des estimateurs de *MCO* en cas d'hétéroscédasticité des erreurs, contrairement aux écarts types estimés calculés par défaut par les logiciels d'économétrie. Bien entendu, la convergence est une propriété de grands échantillons : quand ils sont petits, le résultat peut être médiocre.

5.2 Estimateurs de variance robustes à l'hétéroscédasticité et à l'autocorrélation

Si le terme d'erreur u est autocorrélé, la vraie matrice $\Sigma_{\hat{\beta}MCO|X}$ de variance et de covariance des estimateurs de MCO n'est pas égale à $\sigma_u^2 \left(X'X\right)^{-1}$ et ne peut donc être estimée par $\hat{\Sigma}_{\hat{\beta}MCO|X} = \dfrac{e'e}{n-k}\left(X'X\right)^{-1}$, qui est la formule utilisée par défaut par les logiciels d'économétrie.

Le plus fréquemment, la forme de l'autocorrélation est inconnue, tout comme la formule précise de $\Sigma_{\hat{\beta}MCO|X}$. Il est donc impossible d'établir un estimateur sur mesure de cette matrice, d'autant plus que l'autocorrélation s'accompagne souvent d'hétéroscédasticité, dont la forme est également inconnue. Newey et West [NEW 1987] ont toutefois proposé un estimateur convergent de $\Sigma_{\hat{\beta}MCO|X}$ quelles que soient les formes de l'autocorrélation et de l'hétéroscédasticité. On peut lui appliquer la correction de petit échantillon de MacKinnon et White [MAC 1985]. On obtient la formule [(1)]

$$\hat{\Sigma}^{NW^{adj}}_{\hat{\beta}MCO|X} = \frac{n}{n-k}\left(X'X\right)^{-1}\left(\hat{\Omega}_0 + \sum_{j=1}^{m} w(j,m)\left(\hat{\Omega}_j + \hat{\Omega}'_j\right)\right)\left(X'X\right)^{-1} \qquad (3.23)$$

où $\hat{\Omega}_j = \sum_{t=j+1}^{n} e_t e_{t-j} x_t x'_{t-j}$ pour tout $j = 0 \ldots m$ et où les coefficients $w(j,m)$ déterminent une fenêtre de retards (tronquée à m retards). On distingue plusieurs formes de fenêtre :

- la fenêtre uniforme :

$$w(j,m) = 1 \quad \text{pour tout } \; j = 1 \ldots m$$

- la fenêtre de Bartlett :

$$w(j,m) = 1 - \frac{j}{m+1} \quad \text{pour tout } \; j = 1 \ldots m$$

- la fenêtre de Parzen :

$$w(j,m) = 1 - 6\left(\frac{j}{m+1}\right)^2 + 6\left(\frac{j}{m+1}\right)^3 \quad \text{si } \; 1 \leqslant j \leqslant \frac{m+1}{2}$$

et

$$w(j,m) = 2\left(1 - \frac{j}{m+1}\right)^2 \quad \text{si } \; m \geqslant j > \frac{m+1}{2}$$

1. La formule initiale de Newey et West est $\hat{\Sigma}^{NW}_{\hat{\beta}MCO|X} = \left(X'X\right)^{-1}\left(\hat{\Omega}_0 + \sum_{j=1}^{m} w(j,m)\left(\hat{\Omega}_j + \hat{\Omega}'_j\right)\right)\left(X'X\right)^{-1}$.

La valeur de m doit être choisie arbitrairement. Un terme d'erreur autocorrélé en moyenne mobile d'un ordre connu impose l'usage d'une fenêtre uniforme. C'est le cas en finance, lorsqu'on veut tester l'hypothèse d'efficience d'un marché et que l'horizon des anticipations dépasse la longueur de l'intervalle temporel entre les données successives. Dans les autres cas, la fenêtre de Parzen est généralement préférable aux autres formes. Les fenêtres de Bartlett et de Parzen garantissent l'obtention d'une matrice de variance et de covariance estimée qui soit semi-définie positive, même si m est grand par rapport à n.

Les racines carrées des éléments de la diagonale principale de la matrice de Newey et West sont donc des estimateurs convergents des écarts types des estimateurs de *MCO* en cas d'autocorrélation ou d'hétéroscédasticité des erreurs, contrairement aux écarts types estimés calculés par défaut par les logiciels d'économétrie. Bien entendu, la convergence est une propriété de grands échantillons : quand ils sont petits, le résultat peut être médiocre.

La matrice de White décrite précédemment correspond en fait à celle de Newey et West quand $m = 0$.

6 Estimateur des moindres carrés généralisés

En cas d'autocorrélation ou d'hétéroscédasticité du terme d'erreur u d'un modèle linéaire $Y = X\beta + u$, u n'est pas un bruit blanc : $\Sigma_u \neq \sigma_u^2 I_n$ mais $\Sigma_u = \sigma^2\Omega$ où Ω est définie positive. Il en résulte que, même si le terme d'erreur est indépendant des variables explicatives, la matrice de variance et de covariance de l'estimateur des moindres carrés ordinaires n'est plus égale à $\sigma_u^2 \left(X'X\right)^{-1}$ et ne peut donc être estimée par $\dfrac{e'e}{n-k}\left(X'X\right)^{-1}$. Les écarts types des coefficients de *MCO* publiés par défaut par les logiciels sont donc faux, puisqu'ils sont calculés d'après la formule $\dfrac{e'e}{n-k}\left(X'X\right)^{-1}$, valable uniquement si u est un bruit blanc.

Des estimateurs linéaires sans biais, plus précis que l'estimateur de *MCO* sont disponibles, et celui-ci n'est plus un estimateur linéaire sans biais de variance minimale.

Cela dit, si u est indépendant des variables explicatives, les estimateurs de *MCO* restent sans biais et convergents. Leur précision est en fait fortement altérée par l'autocorrélation et l'hétéroscédasticité. Ces dernières augmentent la probabilité que, sur un échantillon particulier, une valeur estimée soit éloignée de la valeur vraie du coefficient.

Si $Y = X\beta + u$, que $\Sigma_u = \sigma^2\Omega$, que $E(u) = 0$, et si u est indépendant des X_i et Ω connu (σ^2 étant un paramètre inconnu, tout comme le vecteur β), la méthode des moindres carrés généralisés (MCG) permet de mieux estimer β que celle des *MCO*.

L'estimateur de MCG est défini par :

$$\hat{\beta}^{MCG} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \qquad (3.24)$$

Si l'on pose $e = Y - X\beta^{MCG}$, l'expression $\dfrac{e'e}{n-k}$ est un estimateur sans biais de σ^2. On montre aussi que sous les hypothèses décrites précédemment :

$$\Sigma_{\hat{\beta}^{MCG}} = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1} \qquad (3.25)$$

L'application pratique directe des estimateurs des moindres carrés généralisés est assez limitée car elle exige la connaissance préalable de la matrice de variance et de covariance Σ_u du vecteur u des termes d'erreur successifs.

7 Prévision

Après avoir estimé un modèle linéaire, on peut l'utiliser pour produire des prévisions de la variable dépendante sur des périodes postérieures à celles de l'échantillon ayant servi à l'estimation. Ces prévisions sont évidemment conditionnelles à des valeurs supposées des variables explicatives durant les périodes concernées. De manière générale, si le modèle linéaire a été estimé, la prévision ponctuelle de la valeur de la variable dépendante à la période $n + l$ est donnée par :

$$\hat{Y}_{n+l} = \hat{\beta}_1^{MCO} + \hat{\beta}_2^{MCO} X_{2,n+l} + \cdots + \hat{\beta}_k^{MCO} X_{k,n+l} = \hat{\beta}^{MCO'} X_{n+l} \qquad (3.26)$$

où $X_{n+l} = (1 X_{2,n+l} \ldots X_{kn+l})'$. On a obtenu les coefficients estimés en utilisant les matrices Y et X qui rassemblent les données des périodes 1 à n. Sous l'hypothèse que le terme d'erreur u est un bruit blanc et normal, un intervalle de confiance de la valeur de Y_{n+l} est fourni par :

$$P\left(\hat{\beta}^{MCO'} X_{n+l} - t_{\frac{\alpha}{2}} \sqrt{\left(\frac{e'e}{n-k}\right)\left(1 + X'_{n+l}(X'X)^{-1} X_{n+l}\right)} \leqslant Y_{n+l} \leqslant \hat{\beta}^{MCO'} X_{n+l} + t_{\frac{\alpha}{2}} \sqrt{\left(\frac{e'e}{n-k}\right)\left(1 + X'_{n+l}(X'X)^{-1} X_{n+l}\right)}\right) = 1 - \alpha$$

où $t_{\alpha/2}$ est la valeur critique d'une loi de Student à $n - k$ degrés de liberté au seuil de signification $\alpha/2$ et où e est le vecteur des résidus issus de l'estimation du modèle linéaire sur les périodes 1 à n.

8 Modèles linéaires en coupe instantanée

8.1 Généralités

Le concept de modèle linéaire peut s'appliquer à des variables dépendante et explicatives qui n'ont pas la dimension du temps, mais pour lesquelles les différentes observations correspondent à des agents distincts durant une même période de temps. Un modèle linéaire suppose alors l'existence de coefficients non aléatoires $\beta_1, \beta_2 \ldots \beta_k$ identiques quel que soit l'agent concerné, tels que :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{pour tout} \quad i = 1 \ldots n \qquad (3.27)$$

où n est l'effectif des agents différents.

Dans une régression en coupe instantanée, le concept d'autocorrélation, désignant une dépendance du terme d'erreur vis-à-vis de son passé, n'a plus aucun sens tel quel. Les tests d'autocorrélation comme celui de Durbin et Watson ne sont donc pas appliqués en coupe instantanée. Mais cela n'écarte pas le risque que les erreurs de certains agents soient liées entre elles. Le concept d'autocorrélation est simplement remplacé par l'absence d'indépendance entre les erreurs d'agents différents. En série temporelle, l'oubli de variables explicatives importantes, comme une variable dépendante retardée, peut impliquer une dépendance entre les erreurs d'une période particulière t et des périodes précédentes, comme $t - 1$. En coupe instantanée, si l'étude porte sur des agents pouvant être répartis en différentes classes, l'erreur d'un agent particulier d'une classe quelconque risque d'être liée aux erreurs des agents de cette même classe.

8.2 Variables indicatrices et méthodes ANOVA

Comme en série temporelle, on peut utiliser dans des régressions en coupe instantanée des variables indicatrices qui valent, pour chaque agent, 1 ou 0 selon que l'agent concerné appartient ou pas à une certaine catégorie. Le coefficient d'une telle variable exprime alors dans quelle mesure l'appartenance de l'agent à cette catégorie influe sur la valeur de la variable dépendante. Cette analyse est très répandue en sciences de gestion, où elle est connue sous le nom d'ANOVA (ANalysis Of VAriance, en français « analyse de variance »). On distingue plusieurs types d'ANOVA : celle à un facteur (la plus simple) ou encore celle à deux facteurs avec répétition d'expérience. L'objectif général d'une ANOVA est de déterminer si la valeur d'une variable métrique prise (par une personne, une institution, pays, région, entreprise...) dépend des modalités prises par une ou plusieurs caractéristiques qualitatives de cet individu (plus précisément, par une caractéristique dans le cas de l'ANOVA à un facteur, et par deux dans le cas de l'ANOVA à deux facteurs avec répétition d'expérience). Une ANOVA à un facteur revient statistiquement à tester si l'espérance de la variable métrique est égale pour des individus ayant une même modalité de la caractéristique qualitative et pour d'autres individus partageant aussi une même modalité, mais différente de la précédente. Pour l'ANOVA à deux facteurs avec répétition d'expérience, on doit disposer, pour chaque modalité de la première caractéristique qualitative, de groupes d'observations différents de la variable pour chaque modalité différente de la deuxième caractéristique. Vérifier l'hypothèse nulle revient ici à tester si l'espérance de la variable pour un individu pris au hasard est la même quelles que soient :

- la modalité qu'il a pour la caractéristique 1 ;
- la modalité qu'il a pour la caractéristique 2 ;
- la combinaison de modalités qu'il a pour les caractéristiques 1 et 2 (interaction).

9 Méthode des variables instrumentales

9.1 Principe de la méthode

La méthode des variables instrumentales, ou auxiliaires, s'applique lorsque le terme d'erreur d'une équation linéaire n'est pas indépendant de certaine(s) variable(s) explicative(s). Dans ce cas en effet, l'estimateur des moindres carrés ordinaires est biaisé et n'est pas convergent (le biais ne tend donc pas à disparaître si la taille de l'échantillon est grande). Il produirait donc une erreur **systématique** des valeurs estimées. Le cas le plus fréquent où le terme d'erreur n'est pas indépendant de certaine(s) variable(s) explicative(s) est le cas où il y a simultanéité, et donc un impact de la variable dépendante sur une variable explicative.

Soit un modèle linéaire habituel :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots \beta_k X_{kt} + u_t \quad \text{ou} \quad Y = X\beta + u \tag{3.28}$$

où u_t est bruit blanc et où certaine(s) variable(s) explicative(s), parmi les X_i, ne sont pas indépendantes du terme d'erreur u_t. On ne peut donc pas estimer les β_i en calculant par la méthode des moindres carrés ordinaires une régression linéaire des données de Y sur les données des X_i.

L'idée est alors de « purger » les variables explicatives X_i de leur partie liée linéairement à u, c'est-à-dire de remplacer les X_i par des variables transformées $\hat{X}_i$ proches des X_i initiales, mais non liées à u. Bien entendu, on ne touche pas aux variables explicatives X_j, parmi les X_i, qui ne semblent pas liées à u. Pour ces quelques variables X_j, $\hat{X}_j = X_j$.

Comment fabriquer les transformations $\hat{X}_i$? Pour chaque variable explicative X_i, on prend simplement la partie expliquée d'une régression linéaire de X_i sur un ensemble de p variables instrumentales (ou auxiliaires) que l'on suppose de façon arbitraire non liées linéairement à u. Forcément, la partie expliquée de X_i par ces variables instrumentales n'est pas liée à u puisqu'il s'agit d'une fonction linéaire de variables non liées à u. On sélectionne donc *a priori* p variables instrumentales W_r, pour $r = 1 \ldots p$. W_1 est une simple colonne de 1 et donc égale à X_1. Les variables W_r peuvent contenir certaines variables X_i originales : celles supposées indépendantes de u. Il faut que p soit plus grand ou égal à k. Les p colonnes de la matrice W, qui contient n lignes, sont les différentes variables instrumentales W_r, pour $r = 1 \ldots p$.

Pour chaque variable X_i, on calcule par *MCO* le modèle estimé suivant :

$$X_{it} = \gamma_1^{MCO} + \gamma_2^{MCO} W_{2t} + \cdots \gamma_p^{MCO} W_{pt} + e_{it} \tag{3.29}$$

Cela permet de définir la partie expliquée de X_i par les variables instrumentales :

$$\hat{X}_{it} = \hat{\gamma}_{i1}^{MCO} + \hat{\gamma}_{i2}^{MCO} W_{2t} + \cdots \hat{\gamma}_{ip}^{MCO} W_{pt} \tag{3.30}$$

Bien entendu, dans les cas particuliers où X_i figure parmi les variables instrumentales, les coefficients estimés sont tous nuls, sauf celui de la variable elle-même, qui est égal à 1 ; on obtient par ailleurs un résidu toujours nul, ce qui implique que $\hat{X}_{it} = X_{it}$. On estime alors par *MCO* une régression de Y sur les $\hat{X}_i$:

$$Y_t = \hat{\varphi}_1^{MCO} + \hat{\varphi}_2^{MCO} \hat{X}_{2t} + \cdots \hat{\varphi}_k^{MCO} \hat{X}_{kt} + e_t \tag{3.31}$$

Les coefficients estimés obtenus sont les estimateurs de variables instrumentales des coefficients « vrais inconnus » β_i :

$$\hat{\beta}^{VI} = \begin{pmatrix} \beta_1^{VI} \\ \hat{\beta}_2^{VI} \\ \vdots \\ \hat{\beta}_k^{VI} \end{pmatrix} = \begin{pmatrix} \hat{\varphi}_1^{MCO} \\ \hat{\varphi}_2^{MCO} \\ \vdots \\ \hat{\varphi}_k^{MCO} \end{pmatrix} \tag{3.32}$$

Si l'on regroupe dans une matrice X les k variables X_i, X a n lignes et k colonnes. Si l'on regroupe dans une matrice W les p variables W_i, W a n lignes et p colonnes. L'ensemble des régressions de chaque variable X_i sur les variables instrumentales peut donc être représenté de manière matricielle :

$$X = W\hat{\gamma}^{MCO} + E \tag{3.33}$$

où $\hat{\gamma}^{MCO}$ est une matrice à p lignes et k colonnes, chaque colonne étant l'un des vecteurs $\hat{\gamma}_i^{MCO}$ qui regroupe les p coefficients $\hat{\gamma}_{ir}^{MCO}$ de la régression de X_i sur les variables instrumentales W_r, pour $r = 1 \ldots p$. E est une matrice à n lignes et à k colonnes correspondant aux k vecteurs de résidus e_i. On utilise la définition des estimateurs de *MCO* :

$$\hat{\gamma}^{MCO} = (W'W)^{-1} W'X \tag{3.34}$$

On regroupe les k variables $\hat{X}_i$ dans une matrice $\hat{X}$ à n lignes et k colonnes :

$$\hat{X} = W\hat{\gamma}^{MCO} = W(W'W)^{-1}W'X \qquad (3.35)$$

L'estimateur de variables instrumentales de β est alors l'estimateur de MCO de la régression de Y sur les variables $\hat{X}_i$ de $\hat{X}$. Par conséquent, $\hat{\beta}^{VI} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y$, ce qui implique :

$$\hat{\beta}^{VI} = \left(X'W(W'W)^{-1}W'X\right)^{-1} X'W(W'W)^{-1}W'Y \qquad (3.36)$$

Il s'agit de la formule générale de l'estimateur de variables instrumentales. Dans le cas particulier où $p = k$, $W'X$ est une matrice carrée inversible et l'estimateur s'écrit ainsi :

$$\hat{\beta}^{VI} = (W'X)^{-1}W'Y \qquad (3.37)$$

Dans tous les cas, dès lors que les variables instrumentales sont indépendantes de u, on montre que ces estimateurs sont convergents :

$$p\lim \hat{\beta}^{VI} = \beta$$

Les résidus d'une estimation par variables instrumentales sont définis par $e^{VI} = Y - X\hat{\beta}^{VI}$ et un estimateur convergent de σ_u^2 est fourni par la formule $e^{VI'}e^{VI}/(n-k)$ si u est bruit blanc. Un estimateur convergent de la matrice de variance et de covariance de $\hat{\beta}^{VI}$ est alors donné par $\left((e^{VI'}e^{VI})/(n-k)\right)\left(X'W(W'W)^{-1}W'X\right)^{-1}$.

Remarque

Il ne faut pas confondre les résidus d'une estimation par variables instrumentales avec les résidus d'une estimation par moindres carrés ordinaires en deux étapes, ou doubles moindres carrés, définis par $e^{2MC} = Y - \hat{X}\hat{\beta}^{VI}$.

9.2 Qualité de l'ajustement et tests

Les statistiques R^2 et R^2 ajusté, quoique fournies par la plupart des logiciels dans les tableaux de résultats obtenus par variables instrumentales, ne sont en fait pas valables et peuvent conduire à des conclusions erronées [PES 1994]. Dans le contexte des variables instrumentales, on doit leur préférer les statistiques R^2 et R^2 ajustée généralisées. Celles-ci sont calculées d'après les mêmes formules que les statistiques R^2 et R^2 ajustée ordinaires, mais on remplace la série de résidus e par la série $e^{VI} + (X - \hat{X})\hat{\beta}^{VI}$.

Pour vérifier l'hypothèse d'absence d'autocorrélation des erreurs, on ne peut se servir du test de Durbin et Watson car il n'est pas valable sur des résidus estimés par variables instrumentales. On peut par contre utiliser le test LM d'autocorrélation des résidus estimés par variables instrumentales de Sargan, qui est distribué comme une χ^2 à p degrés de liberté sous l'hypothèse nulle d'absence d'autocorrélation du terme d'erreur contre l'hypothèse d'autocorrélation d'ordre p. Ce test est décrit par Breusch et Godfrey [BRE 1981]. Pesaran et Taylor [PES 1997] présentent en détail les tests applicables dans le contexte des variables instrumentales.

Résumé

Dès lors que le terme d'erreur est un bruit blanc et normal, on dispose de la distribution exacte de plusieurs tests qui permettent de vérifier des hypothèses portant simultanément sur plusieurs coefficients d'un modèle linéaire. Sous l'hypothèse de normalité, un modèle linéaire peut aussi être estimé par la méthode du maximum de vraisemblance, qui fournit les mêmes estimateurs des coefficients que ceux obtenus avec les moindres carrés ordinaires. La multicolinéarité réduit la précision des estimateurs de MCO. L'utilisation de variables indicatrices en séries temporelles permet de modéliser la saisonnalité et des événements particuliers non pris en compte par les autres variables explicatives. Alors que les écarts types obtenus de manière classique ne sont corrects que si les erreurs sont un bruit blanc, on dispose d'estimateurs des écarts types des coefficients estimés qui sont robustes même en cas d'hétéroscédasticité ou d'autocorrélation. C'est aussi dans ces situations que les estimateurs de MCG sont plus précis que les estimateurs de MCO. On peut utiliser un modèle estimé par *MCO* pour générer des prévisions ponctuelles ou des intervalles de confiance pour la variable dépendante. La méthode des *MCO* peut s'appliquer à des modèles linéaires reliant des données en coupe instantanée ; dans ce contexte, l'utilisation de certaines variables indicatrices produit des analyses ANOVA. Lorsque le terme d'erreur est lié à certaines variables explicatives, il faut recourir à la méthode des variables instrumentales pour obtenir des estimateurs convergents.



Problèmes et exercices

EXERCICE 1 VENTES D'UNE ENTREPRISE DE GRANDE DISTRIBUTION

Énoncé

Vous disposez de données historiques sur les ventes de l'une des principales chaînes de grande distribution des États-Unis et sur le revenu disponible des ménages américains, dans le fichier de type tableur DISTRIB.xls téléchargeable sur le site Internet www.pearsoneducation.fr. Les définitions des variables de ce fichier sont les suivantes :

VENTES : les ventes totales de tous les magasins de l'entreprise ;

REVDISMN : le revenu disponible des ménages aux États-Unis.

Sur la base de ces données, répondez aux questions suivantes :

- Les achats des ménages dans les magasins de cette entreprise augmentent-ils plus vite, aussi vite ou moins vite que le revenu disponible des ménages ?
- Comment chiffrer à l'avance la croissance des ventes de l'entreprise résultant d'une croissance donnée du revenu disponible des ménages ?

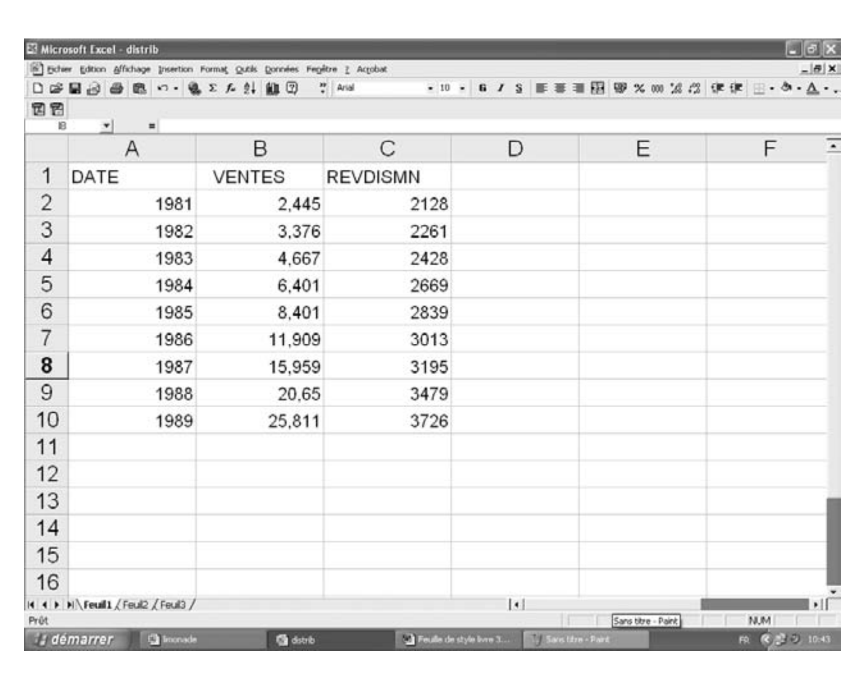
Résolvez ce problème avec un logiciel tableur et TSP.

Solution

Solution avec Excel

Les données se présentent de la manière suivante, dans le fichier DISTRIB.XLS (voir figure 3.1).

Figure 3.1



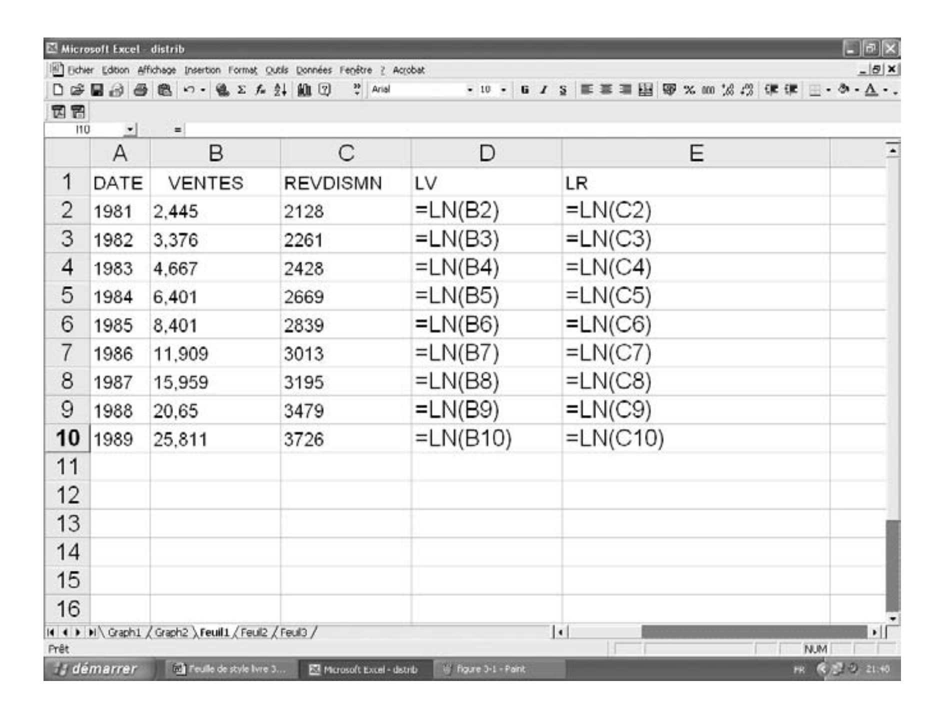
	A	B	C
1	DATE	VENTES	REVDISMN
2	1981	2,445	2128
3	1982	3,376	2261
4	1983	4,667	2428
5	1984	6,401	2669
6	1985	8,401	2839
7	1986	11,909	3013
8	1987	15,959	3195
9	1988	20,65	3479
10	1989	25,811	3726

Les questions de l'exercice portent sur la relation entre les ventes de l'entreprise et le revenu disponible des ménages américains. Pour mesurer cette relation, il est utile de visualiser les données au moyen d'un graphique.

Une relation existe clairement entre les deux variables : les ventes augmentent quand le revenu disponible augmente. Elle ne semble pas tout à fait linéaire, mais plutôt légèrement exponentielle. Il est donc intéressant de visualiser aussi le graphique des logarithmes de ces variables. Pour cela, créez des variables LV et LR égales aux logarithmes népériens des variables initiales (voir figure 3.2).

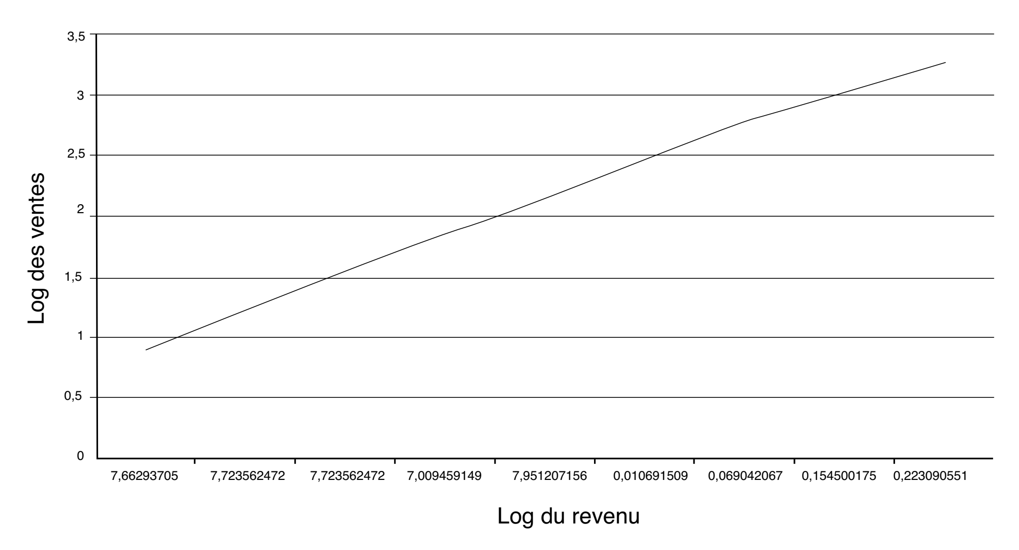
Vous obtenez le graphique de LV en fonction de LR (voir figure 3.3).

Figure 3.2



	A	B	C	D	E
1	DATE	VENTES	REVDISMN	LV	LR
2	1981	2,445	2128	=LN(B2)	=LN(C2)
3	1982	3,376	2261	=LN(B3)	=LN(C3)
4	1983	4,667	2428	=LN(B4)	=LN(C4)
5	1984	6,401	2669	=LN(B5)	=LN(C5)
6	1985	8,401	2839	=LN(B6)	=LN(C6)
7	1986	11,909	3013	=LN(B7)	=LN(C7)
8	1987	15,959	3195	=LN(B8)	=LN(C8)
9	1988	20,65	3479	=LN(B9)	=LN(C9)
10	1989	25,811	3726	=LN(B10)	=LN(C10)
11					
12					
13					
14					
15					
16					

Figure 3.3



La linéarité est plus nette dans la relation entre les logarithmes des séries que dans la relation entre les séries brutes. Par ailleurs, la problématique de départ porte sur une élasticité :

- Les ventes de l'entreprise augmentent plus (respectivement moins) vite que le revenu si l'élasticité des ventes au revenu est supérieure (respectivement inférieure) à 1.
- La réponse de la croissance des ventes à celle du revenu dépend de l'élasticité des ventes au revenu.

Or, l'élasticité des ventes au revenu est la dérivée partielle du logarithme des ventes au logarithme du revenu. Conclusion : c'est la relation entre les logarithmes des ventes et du revenu qui est intéressante. Il faut donc estimer les coefficients du modèle suivant :

$$\ln(VENTES_t) = \beta_1 + \beta_2 \ln(REVDISMN_t) + u_t$$

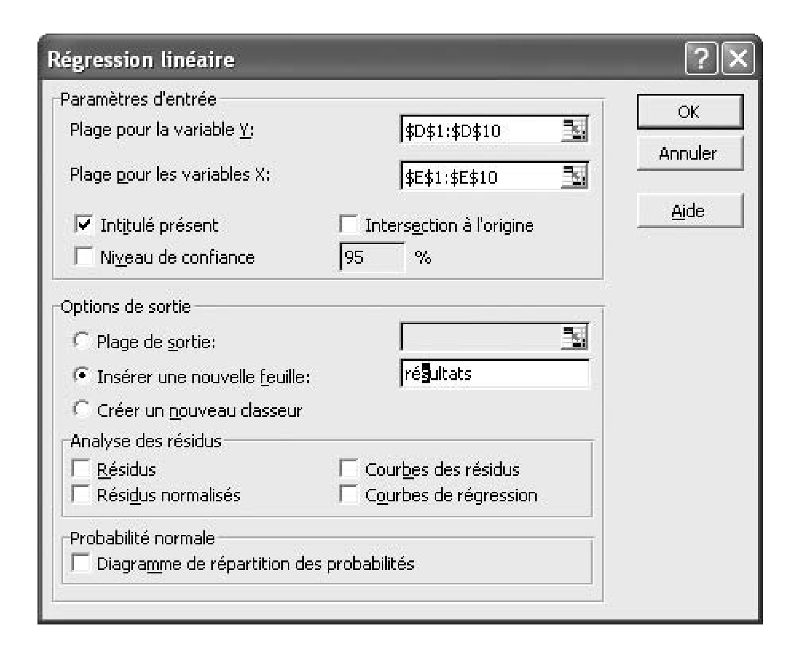
où u_t est le terme d'erreur, c'est-à-dire la partie de l'évolution du logarithme des ventes qui ne s'explique pas linéairement par l'évolution du logarithme du revenu.

Pour réaliser cette estimation par moindres carrés ordinaires avec Excel, procédez ainsi :

Cliquez sur Outils et vérifiez que l'option Utilitaire d'analyse est dans le menu déroulant. Si elle ne s'y trouve pas, cliquez sur l'option Macros complémentaires et, dans le menu déroulant, sélectionnez Utilitaire d'analyse puis cliquez sur OK. Cette fois, au prochain clic sur Outils, vous accéderez à l'Utilitaire d'analyse.

Cliquez donc sur Outils puis sur Utilitaire d'analyse. Dans le menu déroulant, sélectionnez Régression linéaire et cliquez sur OK. Dans Plage pour la variable Y, placez les cellules des données de la variable dépendante : $D$1:$D$10. Dans Plage pour la variable X, placez les cellules des données de la variable explicative : $E$1:$E$10. Cochez l'option Intitulé présent pour indiquer que ces plages contiennent les noms des variables (voir figure 3.4).

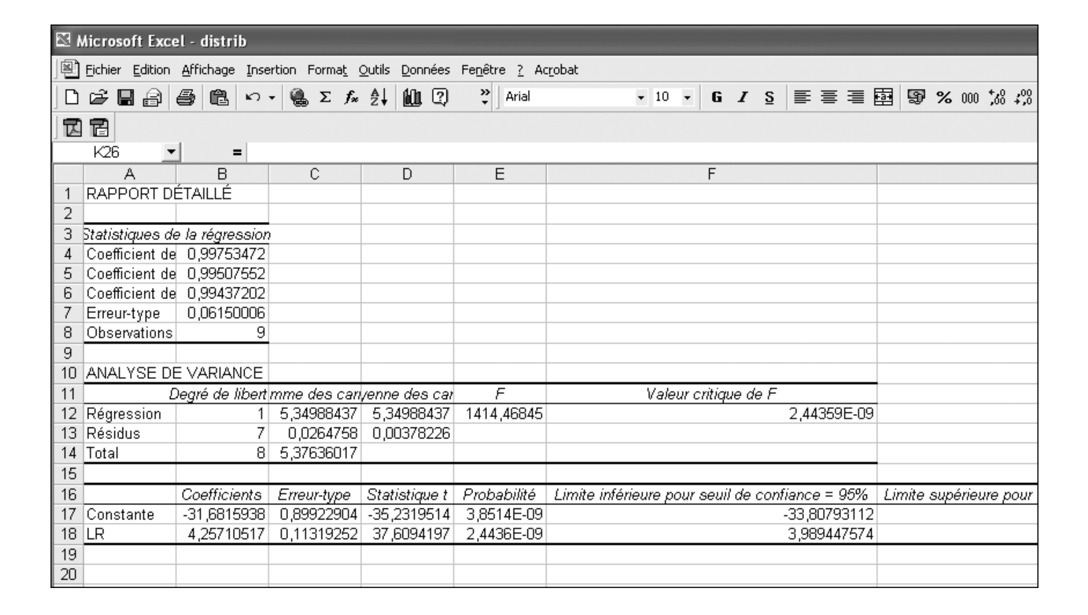
Figure 3.4



Cliquez sur OK. Vous obtenez les résultats (voir figure 3.5, page suivante).

La valeur estimée de β_2, calculée par la formule de l'estimateur $\hat{\beta}_2^{MCO}$, est donc 4,25711. Il s'agit d'une estimation de l'élasticité β_2 des ventes de l'entreprise au revenu disponible des

Figure 3.5



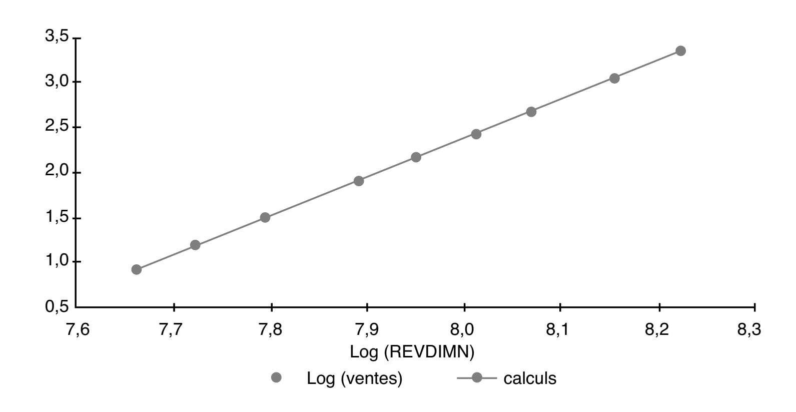
Microsoft Excel - distrib

	A	B	C	D	E	F	
1	RAPPORT DÉTAILLÉ						
2							
3	*Statistiques de la régression*						
4	Coefficient de	0,99753472					
5	Coefficient de	0,99507552					
6	Coefficient de	0,99437202					
7	Erreur-type	0,06150006					
8	Observations	9					
9							
10	ANALYSE DE VARIANCE						
11		*Degré de libert*	*mme des car*	*yenne des car*	*F*	*Valeur critique de F*	
12	Régression	1	5,34988437	5,34988437	1414,46845	2,44359E-09	
13	Résidus	7	0,0264758	0,00378226			
14	Total	8	5,37636017				
15							
16		*Coefficients*	*Erreur-type*	*Statistique t*	*Probabilité*	*Limite inférieure pour seuil de confiance = 95%*	*Limite supérieure pour*
17	Constante	-31,6815938	0,89922904	-35,2319514	3,8514E-09	-33,80793112	
18	LR	4,25710517	0,11319252	37,6094197	2,4436E-09	3,989447574	
19							
20							

ménages, sur la période 1981 à 1989. Par conséquent, de 1981 à 1989, une augmentation de 1 % du revenu disponible des ménages induit en moyenne une augmentation de 4,25711 % des ventes de l'entreprise (voir figure 3.1, page 91).

Le graphique de la figure 3.6, compare les vraies valeurs des ventes en logarithme et les valeurs calculées par la régression ; l'ajustement est excellent. Les erreurs sont très faibles.

Figure 3.6



Réponses aux questions de l'exercice

Durant la période étudiée, les achats des ménages auprès de cette entreprise de grande distribution ont augmenté plus vite que leur revenu disponible puisque l'élasticité estimée des ventes au revenu est supérieure à 1.

En supposant que cette élasticité reste stable dans le temps, vous pouvez l'utiliser pour prévoir la croissance des ventes de l'entreprise conditionnellement à une prévision donnée de croissance du revenu disponible des ménages. Par exemple, si vous prévoyez une croissance de 1,5 % du revenu disponible des ménages, vous pouvez en inférer une prévision de croissance de $4{,}25711 \times 1{,}5 = 6{,}385665$ % des ventes de l'entreprise.



Quelle est la fiabilité de ces conclusions ?

D'après les résultats précédents, l'élasticité des ventes au revenu est estimée à 4,25711 durant la période 1980–1989. Cette valeur estimée est-elle fiable ? Pour répondre à cette question, il faut tenir compte des points suivants :

- Les valeurs calculées de $\hat{\beta}_1^{MCO}$ et $\hat{\beta}_2^{MCO}$ ne sont pas les vraies valeurs de β_1 et β_2, qui restent inconnues. En effet, $\hat{\beta}_1^{MCO}$ et $\hat{\beta}_2^{MCO}$ sont des estimateurs.
- La précision de ces estimations est inversement proportionnelle aux écarts types des estimateurs. Il faut donc examiner les écarts types estimés des coefficients pour connaître leur précision.

Les logiciels donnent des estimations de ces écarts types, calculées selon une méthode qui n'est correcte que si le terme d'erreur est un bruit blanc.

Avant de se fier aux écarts types affichés, il faut donc tester deux hypothèses :

- L'absence d'autocorrélation : le terme d'erreur n'est pas lié à son passé.
- L'absence d'hétéroscédasticité : le terme d'erreur a une dispersion constante.

Pour apprécier la fiabilité des conclusions qui ont été tirées, il faut examiner les résultats de ces tests.

Solution avec TSP

Ouvrez le fichier DISTRIB.XLS initial, contenant les données de la variable VENTES et de la variable REVDISMN, tel qu'il était avant les manipulations réalisées avec le tableur. Il n'est pas nécessaire d'insérer une colonne avec les dates ; mais si vous le faites, nommez-la DATE. Pour que le fichier DISTRIB.XLS puisse être lu par TSP, il faut l'enregistrer comme un fichier de type feuille de calcul Excel (quelle que soit la version), et non comme un classeur. Enregistrez ce fichier dans le répertoire C:\ de l'ordinateur, par exemple. Fermez-le à présent pour pouvoir le lire avec TSP. Pour les raisons évoquées dans le chapitre 2, il faut estimer les coefficients du modèle suivant :

$$\ln(VENTES_t) = \beta_1 + \beta_2 \ln(REVDISMN_t) + u_t$$

Pour travailler avec TSP, suivez la procédure expliquée au chapitre 2. Ici le programme d'instructions est le suivant :

```
FREQ A;
SMPL 1981 1989;
READ(FILE='C:\DISTRIB.XLS');
LV=LOG(VENTES);
LR=LOG(REVDISMN);
REGOPT(PVPRINT,STARS) ALL;
NOPLOT;
OLSQ LV C LR;
```

L'instruction `FREQ A;` indique que la fréquence des données est annuelle. L'instruction `SMPL 1981 1989;` définit la période sur laquelle portent les données. L'instruction `READ(FILE='C:\DISTRIB.XLS');` lit les données dans le fichier DISTRIBT.XLS qui se trouve dans C:\. `LV=LOG(VENTES);` et `LR=LOG(REVDISMN);` fabriquent de nouvelles variables `LV` et `LR` égales au logarithme népérien des variables lues. `REGOPT(PVPRINT,STARS)ALL;` demande tous les tests disponibles (`ALL`), avec mention des probabilités critiques (`PVPRINT`) et présence d'étoiles (`STARS`) si l'hypothèse testée

est rejetée. `NOPLOT;` indique qu'aucun graphique sur les résultats de la régression n'est souhaité. `OLSQ LV C LR;` calcule la régression linéaire de `LV` sur une constante et sur `LR` par MCO.

Les résultats sont les suivants :

```
                         Equation 1
                        ============
                Method of estimation = Ordinary Least Squares
Dependent variable : LV
Current sample : 1981 to 1989
Number of observations : 9
 Mean of dependent variable = 2.12910
Std. dev. of dependent var. = .819784
   Sum of squared residuals = .026475
      Variance of residuals = .378217E-02
   Std. error of regression = .061499
                  R-squared = .995076
         Adjusted R-squared = .994372
    Durbin-Watson statistic = 1.47226
  Wald nonlin. AR1 vs. lags = .122628 [.726]
                  ARCH test = .536009 [.464]
                 CuSum test = .361090 [1.00]
               CuSumSq test = .195113 [.733]
                  Chow test = .343988 [.724]
     LR het. test (w/ Chow) = -1.14689 [1.00]
            White het.  test = 1.67129 [.434]
 Jarque-Bera normality test = .619370 [.734]
  F-statistic (zero slopes) = 1414.50 ** [.000]
   Akaike Information Crit. = -2.54645
 Schwarz Bayes. Info. Crit. = -5.34050
 Log of likelihood function = 13.4590
           Estimated    Standard
Variable   Coefficient    Error       t-statistic     P-value
C          -31.6816       .899219     -35.2323     ** [.000]
LR         4.25711        .113191     37.6098      ** [.000]
```

L'interprétation des coefficients estimés est étudiée au chapitre 2. Des tests supplémentaires permettent d'apprécier la fiabilité des résultats. L'hypothèse d'absence d'autocorrélation du terme d'erreur [1] n'est pas rejetée puisque le test de Durbin et Watson vaut 1,47, pas plus que l'hypothèse d'absence d'hétéroscédasticité du terme d'erreur au seuil de 5 % car les probabilités critiques (entre [ ]) des tests White het, LR het, et ARCH test sont toutes supérieures à 0,05. La validité de ces tests asymptotiques est évidemment très incertaine sur un échantillon aussi petit, mais on admet ici que les erreurs sont homoscédastiques, essentiellement pour des raisons pédagogiques. Vous pouvez donc utiliser les écarts types estimés pour apprécier la précision de l'estimation de β_1 et β_2. L'écart type d'un estimateur est une mesure de l'imprécision de l'estimateur. La précision de l'estimation de β2 est très

1. Pour $n = 15$ et $k = 2$, les valeurs critiques sont $dL = 1{,}08$ et $dU = 1{,}36$ au seuil de 5 %. La valeur du test est supérieure à dU et inférieure à $4 - dU$, ce qui permet de ne pas rejeter l'hypothèse d'absence d'autocorrélation.



grande puisque le coefficient estimé vaut 37 fois son écart type. Un intervalle de confiance pour β2 est donc très étroit autour de 4,25711. Les conclusions tirées sont fiables.

La précision d'un estimateur est d'autant plus faible que son écart type est élevé par rapport à la valeur du coefficient. Inversément, la précision d'un estimateur est d'autant plus élevée que son écart type est faible par rapport à la valeur du coefficient. La colonne *t*-statistic donne, pour le coefficient de chaque variable explicative, le rapport entre sa valeur estimée et l'écart type estimé de l'estimateur. La précision avec laquelle un coefficient est estimé est donc d'autant plus grande que sa *t*-stat est élevée. Ici les coefficients sont estimés avec une très grande précision : les valeurs estimées de β_1 et de β_2 valent, pour la première, 35 fois son écart type et, pour la seconde, 37 fois.

Exercice 2 Ventes de boisson au cola

Énoncé

Il s'agit d'établir, début 1981, une prévision des ventes d'une marque de boisson au cola pour les années 1981 à 1990. Les données de l'entreprise dont vous disposez sont les séries suivantes, de 1970 à 1980 :

SALES : les ventes de l'entreprise à prix courants (nominales) ;

PRICE : le prix de vente de boisson au cola de l'entreprise.

Sont également disponibles des données **externes** qui peuvent être utiles à l'analyse, pour expliquer les ventes de l'entreprise :

Y : le revenu disponible *réel* des ménages (c'est-à-dire le revenu disponible nominal divisé par l'indice des prix à la consommation) ;

CT : la consommation des ménages à prix constants ;

POP : la population ;

GNP : le produit national brut à prix courants ;

CPI : l'indice des prix à la consommation toutes catégories ;

CPI-FB : l'indice des prix à la consommation pour la catégorie « food and beverages » (aliments et boissons).

Ces données se trouvent dans la feuille 1 du fichier limonade.xls téléchargeable sur le site Internet www.pearsoneducation.fr.

Résolvez ce problème d'une part en utilisant un tableur, d'autre part en utilisant TSP.

Solution

Méthodologie

Procédez comme suit :

1. Spécifiez un modèle explicatif des ventes de la boisson au cola en fonction de certaines variables explicatives.
2. Estimez les coefficients « vrais inconnus » de ce modèle sur les données disponibles de 1970 à 1980.
3. Posez des hypothèses sur les valeurs des variables explicatives de 1981 à 1990. Conditionnellement à ces valeurs, utilisez le modèle pour calculer les valeurs des ventes de la boisson au cola pour les années 1981 à 1990.

Spécification du modèle

Supposez que les ventes de cette marque de boisson au cola se déterminent de la manière suivante : les quantités vendues par habitant sont fonction du revenu réel des ménages par habitant et du prix relatif de cette boisson par rapport au prix moyen de l'ensemble des biens et des services de consommation. Cette fonction est vraisemblablement linéaire d'un point de vue logarithmique :

$$\ln\left(\frac{SALES_t}{POP_t PRICE_t}\right) = \beta_1 + \beta_2 \ln\left(\frac{Y_t}{POP_t}\right) + \beta_3 \ln\left(\frac{PRICE_t}{CPI_t}\right) + u_t$$

Que se passe-t-il si vous n'utilisez pas le logarithme de ces variables ? Dans ce cas, vous supposez que :

$$\frac{SALES_t}{POP_t PRICE_t} = \beta_1 + \beta_2 \frac{Y_t}{POP_t} + \beta_3 \frac{PRICE_t}{CPI_t} + u_t$$

Le modèle implique alors qu'à prix relatif inchangé, une augmentation de 1 dollar du revenu entraîne une augmentation de $\beta_{2\,\text{dollars}}$ de la consommation de la boisson, quel que soit le montant de départ du revenu. En d'autres termes, si un riche et un pauvre reçoivent un surplus de revenu de 1 000 dollars, ils vont tous deux augmenter leur consommation de ladite boisson d'un même montant. Cela n'a évidemment aucun sens !

Le modèle en logarithmes peut encore être présenté de la manière suivante :

$$LRS_t = \beta_1 + \beta_2 LYP_t + \beta_3 LRPR_t + u_t$$

où $LRS_t = \ln\left(\frac{SALES_t}{POP_t PRICE_t}\right)$, $LY_t = \ln\left(\frac{Y_t}{POP_t}\right)$, $LRPR_t = \ln\left(\frac{PRICE_t}{CPI_t}\right)$.

Estimation du modèle avec un tableur

Le fichier de données se présente comme à la figure 3.7.

Figure 3.7

Microsoft Excel - limonade

	A	B	C	D	E	F	G	H	I
1		Sales	Population	GNP	CPI-FB	Price	CPI	Y	CT
2	1970	1606,4	205	32,3	114,7	9,69	35,647	1975,04	1.813,470
3	1971	1728,8	208	35,4	118,3	9,96	37,376	2050,70	1.873,720
4	1972	1876,2	210	38,6	123,2	10,03	38,809	2132,58	1.978,440
5	1973	2145	212	42,1	139,4	10,31	41,038	2276,86	2.066,740
6	1974	2522,2	214	45,8	158,7	12,88	45,170	2259,70	2.053,810
7	1975	2872,8	216	50,6	172,1	15	48,863	2303,23	2.097,500
8	1976	3032,8	218	55,6	177,4	14,64	51,787	2387,57	2.207,250
9	1977	3559,9	220	60,5	188	14,79	55,366	2455,96	2.296,570
10	1978	4337,7	223	67,8	206,3	15,93	59,419	2574,33	2.391,810
11	1979	4961,4	225	75,6	228,5	17,02	64,685	2639,30	2.448,350
12	1980	5912,6	228	85,3	248	19,3	71,436	2663,03	2.447,070

Commencez par créer dans les colonnes J, K et L les variables $RS_t = \dfrac{SALES_t}{POP_t PRICE_t}$, $Y_t = \dfrac{Y_t}{POP_t}$, $RPR_t = \left(\dfrac{PRICE_t}{CPI_t}\right)$. Les formules se présentent comme à la figure 3.8.

Figure 3.8

	J	K	L
1	RS	YP	RPR
2	=(B2/F2)/C2	=H2/C2	=F2/G2
3	=(B3/F3)/C3	=H3/C3	=F3/G3
4	=(B4/F4)/C4	=H4/C4	=F4/G4
5	=(B5/F5)/C5	=H5/C5	=F5/G5
6	=(B6/F6)/C6	=H6/C6	=F6/G6
7	=(B7/F7)/C7	=H7/C7	=F7/G7
8	=(B8/F8)/C8	=H8/C8	=F8/G8
9	=(B9/F9)/C9	=H9/C9	=F9/G9
10	=(B10/F10)/C10	=H10/C10	=F10/G10
11	=(B11/F11)/C11	=H11/C11	=F11/G11
12	=(B12/F12)/C12	=H12/C12	=F12/G12
13			

Puis calculez les valeurs de LRS, LYP et LRPR dans les colonnes M, N et O (voir figure 3.9).

Figure 3.9

	N	O	P
1	LRS	LYP	LRPRFB
2	=LN(J2)	=LN(K2)	=LN(L2)
3	=LN(J3)	=LN(K3)	=LN(L3)
4	=LN(J4)	=LN(K4)	=LN(L4)
5	=LN(J5)	=LN(K5)	=LN(L5)
6	=LN(J6)	=LN(K6)	=LN(L6)
7	=LN(J7)	=LN(K7)	=LN(L7)
8	=LN(J8)	=LN(K8)	=LN(L8)
9	=LN(J9)	=LN(K9)	=LN(L9)
10	=LN(J10)	=LN(K10)	=LN(L10)
11	=LN(J11)	=LN(K11)	=LN(L11)
12	=LN(J12)	=LN(K12)	=LN(L12)
13			

Pour estimer les coefficients du modèle par MCO, sélectionnez successivement Outils/Utilitaire d'analyse/Régression linéaire et cliquez sur OK. Remplissez certaines cases de l'écran de régression à la figure 3.10, page suivante.

Cliquez sur OK. Les résultats sont identiques à ceux de la figure 3.11, page suivante.

Le logiciel a attribué à la feuille de résultats le nom arbitraire que vous avez choisi de lui donner : OLS. Vous obtenez des valeurs estimées pour les coefficients « vrais inconnus » β_1, β_2 et β_3 : $\hat{\beta}_1^{MCO} = -6{,}84817$, $\hat{\beta}_2^{MCO} = 2{,}492627$, $\hat{\beta}_3^{MCO} = -0{,}71303$.

L'élasticité des quantités vendues par habitant au revenu disponible réel par habitant est mesurée par β_2 et sa valeur estimée vaut 2,492627 : par conséquent, si le revenu augmente

Figure 3.10

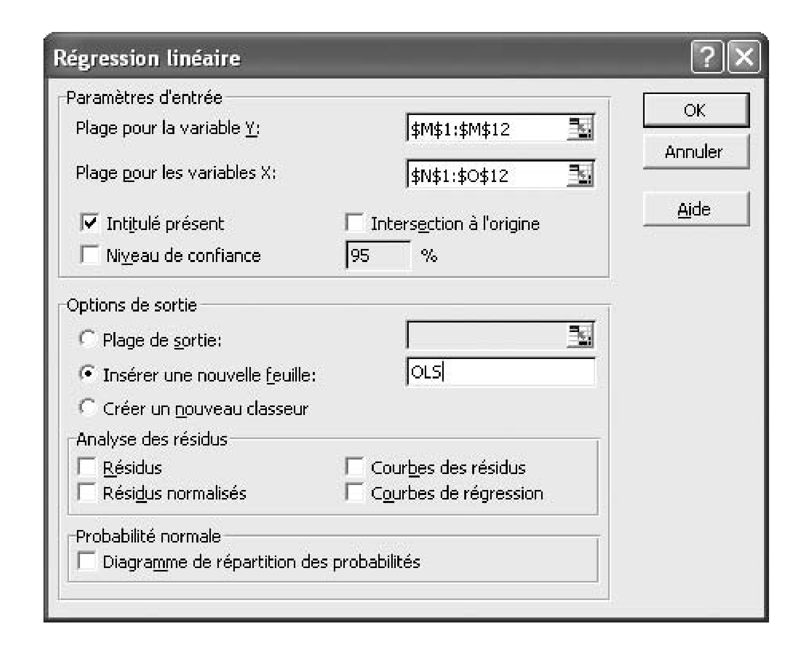
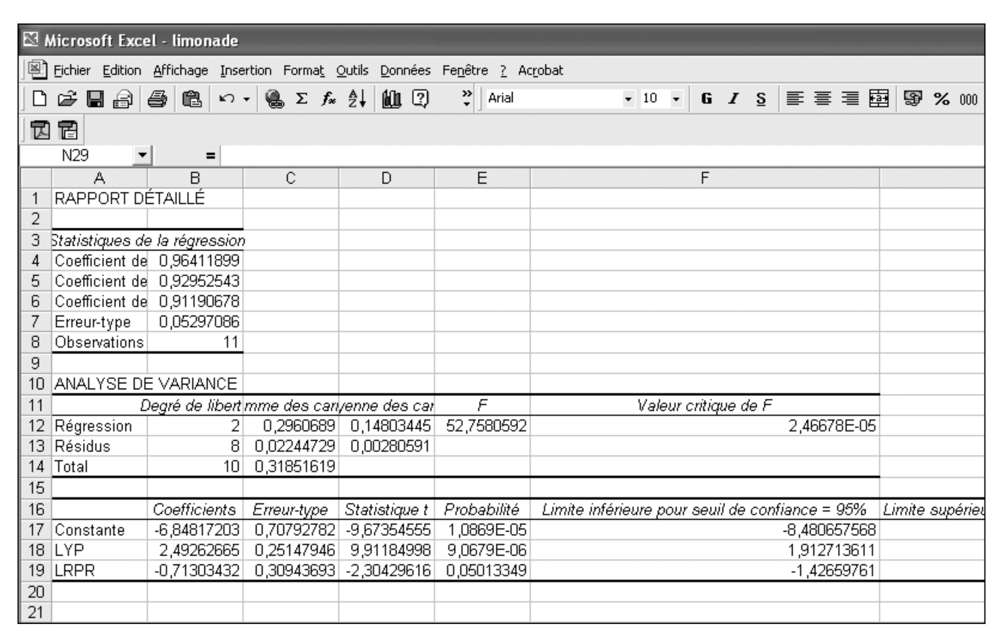


Figure 3.11


Microsoft Excel - limonade

	A	B	C	D	E	F	
1	RAPPORT DÉTAILLÉ						
2							
3	Statistiques de la régression						
4	Coefficient de	0,96411899					
5	Coefficient de	0,92952543					
6	Coefficient de	0,91190678					
7	Erreur-type	0,05297086					
8	Observations	11					
9							
10	ANALYSE DE VARIANCE						
11		Degré de libert	mme des car	yenne des car	F	Valeur critique de F	
12	Régression	2	0,2960689	0,14803445	52,7580592	2,46678E-05	
13	Résidus	8	0,02244729	0,00280591			
14	Total	10	0,31851619				
15							
16		Coefficients	Erreur-type	Statistique t	Probabilité	Limite inférieure pour seuil de confiance = 95%	Limite supérieu
17	Constante	-6,84817203	0,70792782	-9,67354555	1,0869E-05	-8,480657568	
18	LYP	2,49262665	0,25147946	9,91184998	9,0679E-06	1,912713611	
19	LRPR	-0,71303432	0,30943693	-2,30429616	0,05013349	-1,42659761	
20							
21							

de 1 %, les quantités en question augmentent de 2,492627 %. L'élasticité des quantités vendues de boisson au cola par habitant à son prix relatif est mesurée par β_3 et sa valeur estimée vaut −0,71303 : par conséquent, si le prix augmente de 1 %, les quantités en question diminuent de 0,71303 %. R^2 est égal à 0,929525 : en d'autres termes, 92,9525 % des fluctuations des quantités vendues en logarithmes sont expliquées par les fluctuations du revenu disponible réel en logarithme et du prix relatif en logarithme, sur la période 1970–1980.

L'écart type estimé de l'estimateur $\hat{\beta}_1^{MCO}$ est égal à 0,707928, celui de $\hat{\beta}_2^{MCO}$ à 0,251479 et celui de $\hat{\beta}_3^{MCO}$ à 0,309437. Plus l'écart type de l'estimateur d'un coefficient est élevé, plus grande est l'imprécision avec laquelle le coefficient est estimé. Pour évaluer si l'écart type



estimé est grand ou petit, comparez-le à la valeur estimée du coefficient. La colonne t Stat permet cette comparaison puisque t-stat = coefficient estimé/écart type estimé. Donc, plus une t-stat est élevée, plus la précision de l'estimation du coefficient concerné est grande. De manière générale, à partir de 2, la précision est bonne et, en dessous de 1, elle est très insuffisante. Les écarts types estimés par les logiciels ne sont fiables que si le terme d'erreur n'est ni autocorrélé ni hétéroscédastique car ils sont calculés à l'aide d'une formule qui n'est correcte que si le terme d'erreur est un bruit blanc. Par ailleurs, en supposant que ce dernier est distribué normalement, et en l'absence d'autocorrélation et d'hétéroscédasticité, vous ne pouvez rejeter l'hypothèse qu'un coefficient « vrai inconnu » est nul si sa t-stat est inférieure à la valeur critique à 5 % d'une distribution Student à 8 = 11(observations) − 3(coefficients) degrés de liberté. Dans ce cas, la probabilité critique (p-value en anglais) fournit le résultat puisque vous ne rejetez pas l'hypothèse qu'un coefficient « vrai inconnu » est nul si la probabilité critique est supérieure à 0,05. Mais comment savoir si le terme d'erreur n'est ni autocorrélé, ni hétéroscédastique, et s'il est distribué normalement ? Utilisez les tests d'absence d'autocorrélation, d'absence d'hétéroscédasticité et de normalité du terme d'erreur (voir plus loin).

Extrapolation avec un tableur

Vous allez réaliser une extrapolation des valeurs des quantités vendues par l'entreprise de boisson au cola pour les années 1981 à 1990, en utilisant Excel.

Pour cela, il faut d'abord poser les hypothèses suivantes sur l'évolution des variables explicatives, de 1981 à 1990 :

- L'entreprise envisage d'augmenter son prix de vente (PRICE) de 10 % chaque année.
- L'indice des prix à la consommation (CPI) va augmenter de 8 % annuellement. Ces hypothèses supposent que le prix relatif (RPR) augmente de 2 % par an.
- Le revenu disponible réel par habitant (YP) va augmenter de 2 % chaque année.
- La population va augmenter de 1,5 % chaque année.

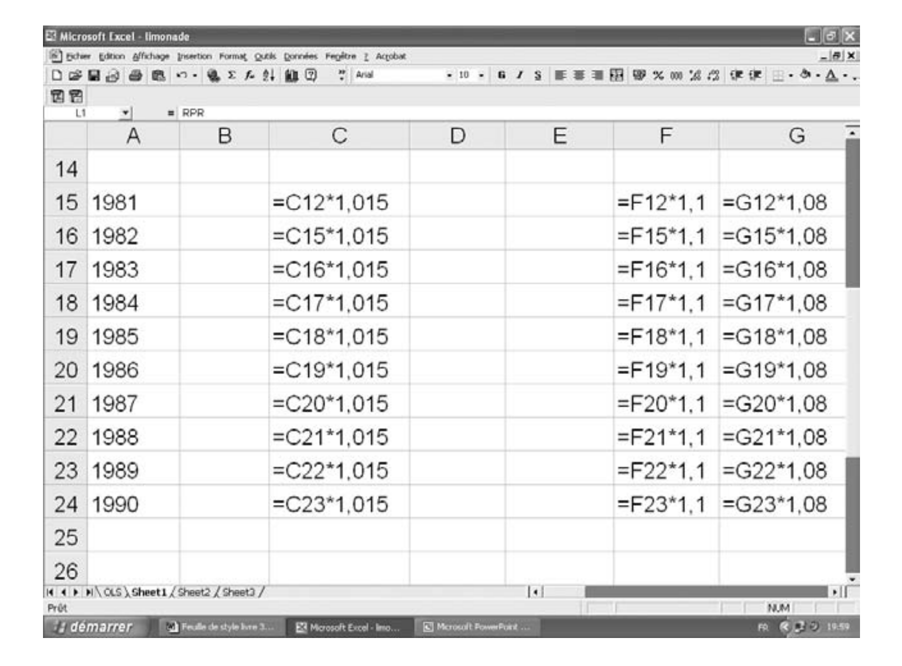
Dans le fichier limonade.xls, affichez la feuille initiale (appelée ici limonade). Prolongez les dates de la colonne A, inscrivez les valeurs supposées de YP, PRICE, CPI pour les années 1981 à 1990 en colonnes G, G et K, et calculez leurs implications sur RPR en colonne L, à partir de la ligne 15. Calculez les valeurs supposées de LYP et LRPR de 1981 à 1990, en colonnes N et O (voir figures 3.12, page suivante et 3.13, page suivante).

Maintenant que les hypothèses sur les variables explicatives, de 1981 à 1990, sont posées, il est possible d'utiliser le modèle estimé pour calculer les valeurs de LRS qui en résultent, pour 1981 à 1990. Les résultats de l'estimation du modèle sont dans la feuille OLS du fichier limonade.xls. En particulier, les coefficients estimés sont donnés par les cellules suivantes de cette feuille : $\hat{\beta}_1^{MCO}$ = OLS!$B$17, $\hat{\beta}_2^{MCO}$ = OLS!$B$18, $\hat{\beta}_3^{MCO}$ = OLS!$B$19. Il faut calculer les prévisions de LRS à l'aide des formules suivantes :

$$LRS_{1981}^{p} = \hat{\beta}_1^{MCO} + \hat{\beta}_2^{MCO} LYP_{1981}^{hyp} + \hat{\beta}_3^{MCO} LRPR_{1981}^{hyp}$$
$$LRS_{1982}^{p} = \hat{\beta}_1^{MCO} + \hat{\beta}_2^{MCO} LYP_{1982}^{hyp} + \hat{\beta}_3^{MCO} LRPR_{1982}^{hyp}$$
$$\ldots$$
$$LRS_{1990}^{p} = \hat{\beta}_1^{MCO} + \hat{\beta}_2^{MCO} LYP_{1990}^{hyp} + \hat{\beta}_3^{MCO} LRPR_{1990}^{hyp}$$

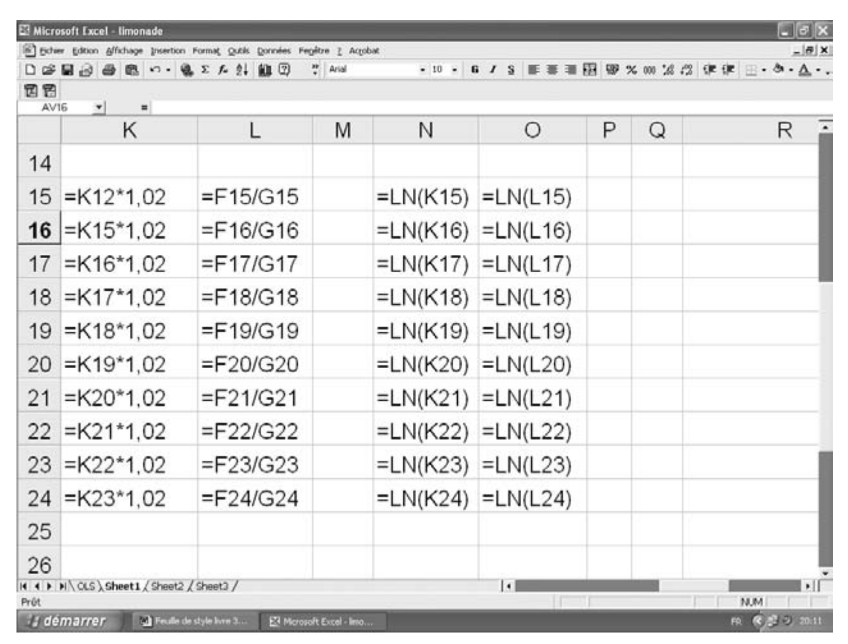
Faites les modifications qui s'imposent dans la feuille limonade du fichier limonade.xls (voir figure 3.14, page 103).

Figure 3.12



	A	B	C	D	E	F	G
14							
15	1981		=C12*1,015			=F12*1,1	=G12*1,08
16	1982		=C15*1,015			=F15*1,1	=G15*1,08
17	1983		=C16*1,015			=F16*1,1	=G16*1,08
18	1984		=C17*1,015			=F17*1,1	=G17*1,08
19	1985		=C18*1,015			=F18*1,1	=G18*1,08
20	1986		=C19*1,015			=F19*1,1	=G19*1,08
21	1987		=C20*1,015			=F20*1,1	=G20*1,08
22	1988		=C21*1,015			=F21*1,1	=G21*1,08
23	1989		=C22*1,015			=F22*1,1	=G22*1,08
24	1990		=C23*1,015			=F23*1,1	=G23*1,08
25							
26							

Figure 3.13



	K	L	M	N	O	P	Q	R
14								
15	=K12*1,02	=F15/G15		=LN(K15)	=LN(L15)			
16	=K15*1,02	=F16/G16		=LN(K16)	=LN(L16)			
17	=K16*1,02	=F17/G17		=LN(K17)	=LN(L17)			
18	=K17*1,02	=F18/G18		=LN(K18)	=LN(L18)			
19	=K18*1,02	=F19/G19		=LN(K19)	=LN(L19)			
20	=K19*1,02	=F20/G20		=LN(K20)	=LN(L20)			
21	=K20*1,02	=F21/G21		=LN(K21)	=LN(L21)			
22	=K21*1,02	=F22/G22		=LN(K22)	=LN(L22)			
23	=K22*1,02	=F23/G23		=LN(K23)	=LN(L23)			
24	=K23*1,02	=F24/G24		=LN(K24)	=LN(L24)			
25								
26								

Il reste à transformer les prévisions de LRS (les cellules M15 à M24) en prévisions pour les ventes nominales totales (SALES) de la manière suivante :

$$SALES^{p}_{1981} = POP_{1981} * PRICE_{1981} * \left(e^{LRS^{p}_{1981}}\right)$$

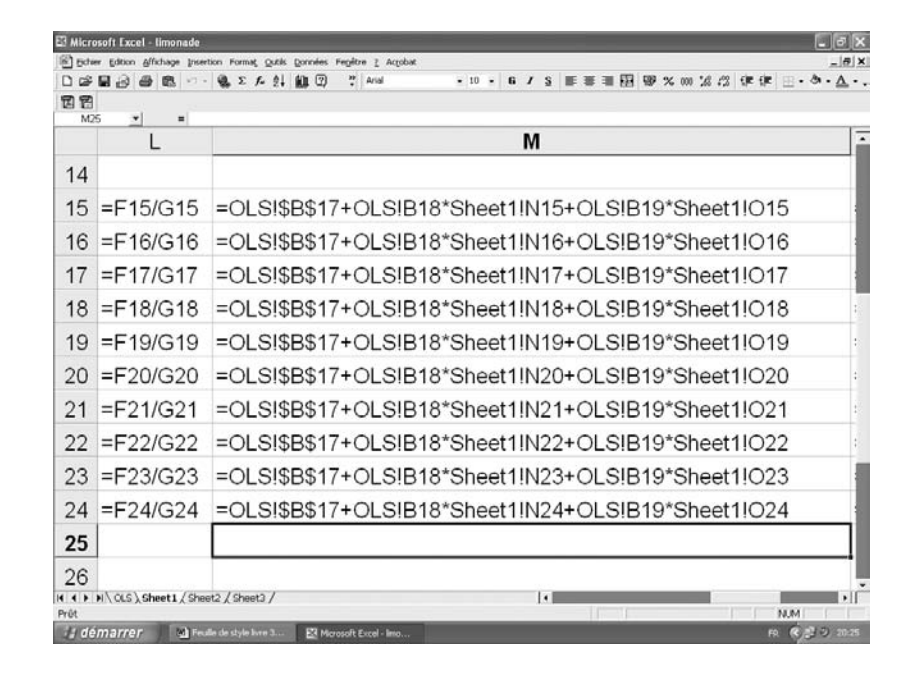
$$SALES^{p}_{1982} = POP_{1982} * PRICE_{1982} * \left(e^{LRS^{p}_{1982}}\right)$$

$$\ldots$$

$$SALES^{p}_{1990} = POP_{1990} * PRICE_{1990} * \left(e^{LRS^{p}_{1990}}\right)$$



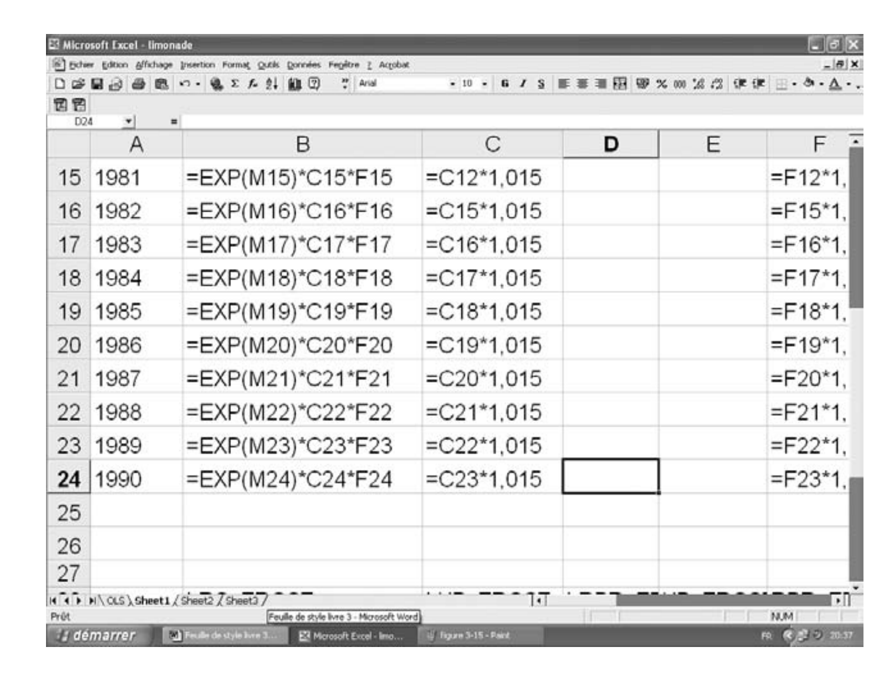
Figure 3.14



	L	M
14		
15	=F15/G15	=OLS!$B$17+OLS!B18*Sheet1!N15+OLS!B19*Sheet1!O15
16	=F16/G16	=OLS!$B$17+OLS!B18*Sheet1!N16+OLS!B19*Sheet1!O16
17	=F17/G17	=OLS!$B$17+OLS!B18*Sheet1!N17+OLS!B19*Sheet1!O17
18	=F18/G18	=OLS!$B$17+OLS!B18*Sheet1!N18+OLS!B19*Sheet1!O18
19	=F19/G19	=OLS!$B$17+OLS!B18*Sheet1!N19+OLS!B19*Sheet1!O19
20	=F20/G20	=OLS!$B$17+OLS!B18*Sheet1!N20+OLS!B19*Sheet1!O20
21	=F21/G21	=OLS!$B$17+OLS!B18*Sheet1!N21+OLS!B19*Sheet1!O21
22	=F22/G22	=OLS!$B$17+OLS!B18*Sheet1!N22+OLS!B19*Sheet1!O22
23	=F23/G23	=OLS!$B$17+OLS!B18*Sheet1!N23+OLS!B19*Sheet1!O23
24	=F24/G24	=OLS!$B$17+OLS!B18*Sheet1!N24+OLS!B19*Sheet1!O24
25		
26		

Pour cela, il faut introduire pour commencer les formules de calcul des valeurs supposées de la population de 1981 à 1990, puis insérer les formules de calcul des prévisions de SALES (voir figure 3.15).

Figure 3.15



	A	B	C	D	E	F
15	1981	=EXP(M15)*C15*F15	=C12*1,015			=F12*1,
16	1982	=EXP(M16)*C16*F16	=C15*1,015			=F15*1,
17	1983	=EXP(M17)*C17*F17	=C16*1,015			=F16*1,
18	1984	=EXP(M18)*C18*F18	=C17*1,015			=F17*1,
19	1985	=EXP(M19)*C19*F19	=C18*1,015			=F18*1,
20	1986	=EXP(M20)*C20*F20	=C19*1,015			=F19*1,
21	1987	=EXP(M21)*C21*F21	=C20*1,015			=F20*1,
22	1988	=EXP(M22)*C22*F22	=C21*1,015			=F21*1,
23	1989	=EXP(M23)*C23*F23	=C22*1,015			=F22*1,
24	1990	=EXP(M24)*C24*F24	=C23*1,015			=F23*1,
25						
26						
27						

Inscrivez dans une cellule surmontant les valeurs extrapolées de SALES la mention SALES-FRCST pour les désigner.

Vous obtenez les valeurs suivantes :

```
6294,69848
7287,66761
```

```
8437,27454
9768,22839
11309,1361
13093,1172
15158,5157
17549,7245
20318,139
23523,2623
```

Estimation et extrapolation avec TSP

Avec TSP, le principe et la procédure sont différents : créez un nouveau fichier Excel, dans lequel vous recopiez certaines données historiques du fichier initial limonade.XLS, en l'occurrence seulement celles dont vous avez besoin dans l'analyse : les dates, les données de SALES, Population, PRICE, CPI et Y. Sont donc exclues les données de GNP, CPI-FB et CT (vous pourriez insérer ces variables dans le fichier, même sans les utiliser par la suite, mais il faudrait alors changer le nom CPI-FB en CPIFB : TSP n'accepte pas les noms de variables contenant des tirets).

Ajoutez l'en-tête DATE au-dessus de la colonne de dates. Changez le nom Population en POP parce que certaines versions anciennes de TSP n'acceptent que des noms de huit caractères maximum. Enregistrez le document comme un fichier de type feuille de calcul Excel (quelle que soit la version), sous le nom LIM (l'ordinateur lui donne automatiquement l'extension .xls), dans le répertoire C:\ de l'ordinateur par exemple. Fermez-le pour pouvoir le lire avec TSP. Pour travailler sous TSP, suivez la procédure décrite au chapitre 2. Ici le programme d'instructions est le suivant :

```
FREQ A;
SMPL 1970 1980;
? LECTURE DES DONNEES
READ(FILE='C :\LIM.XLS');
? CREATION DE VARIABLES
RS=(SALES/POP)/PRICE;
YP=Y/POP;
RPR=PRICE/CPI;
LRS=LOG(RS);
LYP=LOG(YP);
LRPR=LOG(RPR);
? CALCUL DE REGRESSION
REGOPT(PVPRINT,STARS) ALL;
NOPLOT;
OLSQ LRS C LYP LRPR;
? HYPOTHESES EN PREVISION
SMPL 1981 1990;
YP=YP(-1)*1.02;
POP=POP(-1)*1.015;
CPI=CPI(-1)*1.08;
PRICE=PRICE(-1)*1.10;
RPR=PRICE/CPI;
```

```
LYP=LOG(YP);
LRPR=LOG(RPR);
? CALCUL DE PREVISION EN LOG SUR LES VENTES REELLES PAR HABITANT
FORCST(PRINT) LRSP;
? CALCUL DE LA PREVISION SUR LES VENTES NOMINALES TOTALES
SALESP=EXP(LRSP)*POP*PRICE;
PRINT SALESP;
```

Les lignes de code commençant par un point d'interrogation (?) sont des commentaires qui décrivent le programme (vous pouvez donc les ignorer lorsque vous réalisez cet exercice). Les résultats sont les suivants :

```
 Mean of dependent variable = .510351E-02
Std. dev. of dependent var. = .178470
   Sum of squared residuals = .022447
      Variance of residuals = .280592E-02
   Std. error of regression = .052971
                  R-squared = .929525
         Adjusted R-squared = .911907
    Durbin-Watson statistic = .961980
Breusch/Godfrey LM : AR/MA1 = 1.97215 [.160]
Ljung-Box Q-statistic1 = 1.12998 [.288]
  Wald nonlin. AR1 vs. lags = 5.16845 [.075]
                  ARCH test = 1.66231 [.197]
                 CuSum test = .770092 [.166]
               CuSumSq test = .487032 * [.037]
                  Chow test = 8.68049 * [.020]
LR het.  test (w/ Chow) = 26.9613 ** [.000]
White het. test = 1.11500 [.953]
Jarque-Bera normality test = .118607 [.942]
  F-statistic (zero slopes) = 52.7580 ** [.000]
  Akaike Information Crit.  = -2.81115
 Schwarz Bayes. Info. Crit. = -5.54051
 Log of likelihood function = 18.4613
           Estimated    Standard
Variable  Coefficient     Error      t-statistic     P-value
C         -6.84817       .707928     -9.67354     ** [.000]
LYP       2.49263        .251480     9.91184      ** [.000]
LRPR      -.713034       .309437     -2.30429        [.050]
               SALESP
1981       6294.69873
1982       7287.66650
1983       8437.27344
1984       9768.23047
1985      11309.13770
1986      13093.11816
1987      15158.51367
1988      17549.73047
1989      20318.14258
1990      23523.26367
```

Exercices

Vous retrouvez certains résultats déjà obtenus précédemment, mais des tests permettent de mieux apprécier leur validité. Deux des trois tests d'hétéroscédasticité ont une probabilité critique supérieure à 0,05 et ne rejettent donc pas l'hypothèse que le terme d'erreur n'est pas hétéroscédastique. Il s'agit des tests ARCH (qui est peu pertinent ici sur des données annuelles) et du test White het. Les tests de Breusch et Godfrey, de LM : AR/MA1 et de Ljung et Box (Q-statistic1) ont une probabilité critique supérieure à 0,05 et ne rejettent donc pas l'hypothèse que le terme d'erreur n'est pas autocorrélé. Le terme d'erreur semble être un bruit blanc, ce qui permet d'affirmer que les écarts types estimés de la colonne standard errors sont fiables. Les *t*-stats élevées montrent que les coefficients sont estimés d'une manière précise.

EXERCICE 3 LE CAS BANQUE RÉGIONALE FRANÇAISE

Énoncé

Une agence urbaine d'une grande banque régionale française [1] souhaite, en octobre 2001, mettre au point une méthode de prévision de ses volumes mensuels de production de crédits octroyés sous forme de prêts personnels. Cette agence est située dans le centre d'une grande métropole. Les données disponibles en provenance de cette banque sont :

- une série temporelle mensuelle avec le volume observé de production de crédits de l'agence Lille Métropole, de janvier 1999 à septembre 2001 (35 mois) ;
- des séries temporelles mensuelles de l'encours, des rachats et des échus de crédits de cette agence pour les mêmes mois ;
- des séries temporelles mensuelles de la marge sur les crédits produits et de la marge sur les crédits en cours de l'agence pour les mêmes mois.

Ces données sont sur la première feuille du fichier BA.XLS :

- en colonne A, la variable date : les dates successives de novembre 1998 à septembre 2001 ;
- en colonne B, la variable pp_pro : la production de prêts personnels ;
- en colonne C, la variable pp_enc : l'encours des prêts personnels ;
- en colonne D, la variable pp_ra : les rachats de prêts personnels ;
- en colonne E, la variable pp_ec : les échus de prêts personnels ;
- en colonne F, la variable pp_pro_m : la marge sur la production des prêts personnels ;
- en colonne G, la variable pp_enc_m : la marge sur l'encours des prêts personnels.

Pour chaque colonne, le nom de la variable est en ligne 1 ; les trente-cinq observations successives viennent sur les lignes suivantes.

Solution

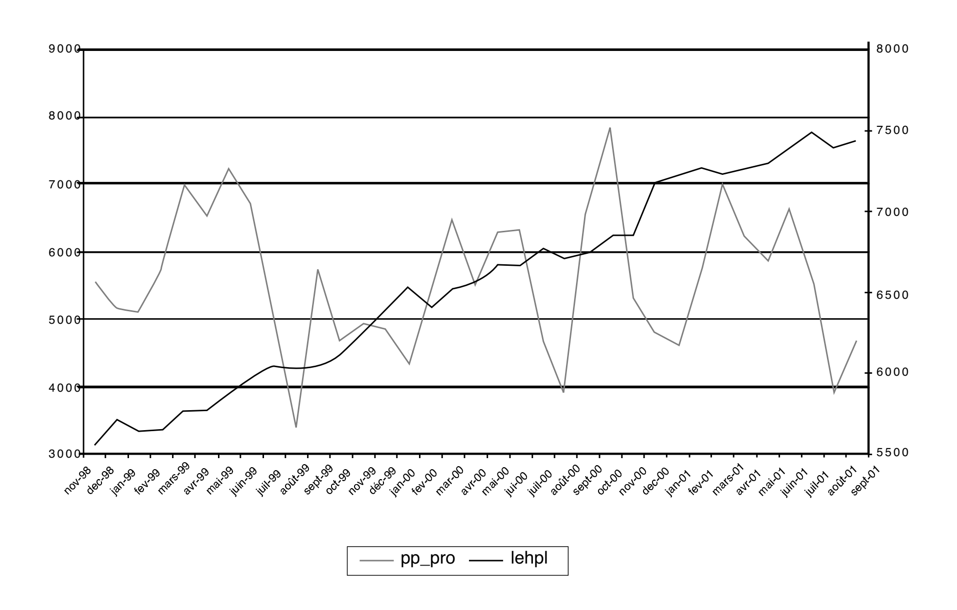
Exploration d'une première piste

Pour élaborer une méthode de prévision de la production de crédits prêts personnels pour l'agence bancaire, il faut d'abord comprendre les facteurs qui peuvent expliquer l'évolution de cette production.

1. Il s'agit d'une demande réelle traitée à l'IESEG pour le compte d'une agence d'une grande banque régionale française. Celle-ci ayant souhaité la confidentialité, son identité ne sera pas divulguée ici. La méthodologie de prévision devait être mise au point pour trois catégories différentes de crédits : les prêts acquéreurs, qui sont des prêts hypothécaires, les prêts personnels et les prêts équipements MLT (Moyen Long Terme), qui sont des prêts accordés à des entreprises et à des artisans pour l'achat de biens d'investissement.

Si les productions de crédits prêts personnels de l'agence et de l'ensemble des banques de France semblent suivre globalement la même évolution, vous pouvez expliquer la production de crédits de l'agence par les mêmes facteurs que ceux qui influencent la production totale de crédits en France. Il s'agit de facteurs nationaux globaux relativement bien connus, faisant l'objet de statistiques publiées par des organismes officiels comme l'INSEE, le FMI, l'OCDE. Pour vérifier que la production de crédits de l'agence concernée suit la tendance nationale, vous allez comparer graphiquement son évolution à celle de la production de crédits prêts personnels de l'ensemble des banques de France. Cette variable nationale figure dans les statistiques financières internationales du FMI : elle s'intitule FR BANK LENDING TO HOUSEHOLDS AS PERSONAL LOANS CUR. Ajoutez ces données à la colonne H du fichier CN.XLS, avec comme nom l'abréviation lehpl. Comparez alors la production de prêts personnels de l'agence avec la production nationale, représentées sur un même graphique (voir figure 3.16).

Figure 3.16



La production de crédits en prêts personnels de l'agence ne suit pas la tendance nationale.

Remarque

Les divergences entre évolution de l'agence et évolution nationale ne viennent pas de la différence entre le caractère lisse de la série nationale et le caractère erratique de la série de l'agence, qui s'explique par le fait que la série nationale est « corrigée des variations saisonnières », ce qui n'est pas le cas de la série de l'agence. Elles sont dues à la tendance croissante affichée par la série nationale (et non par la série de l'agence) tout au long de la période.

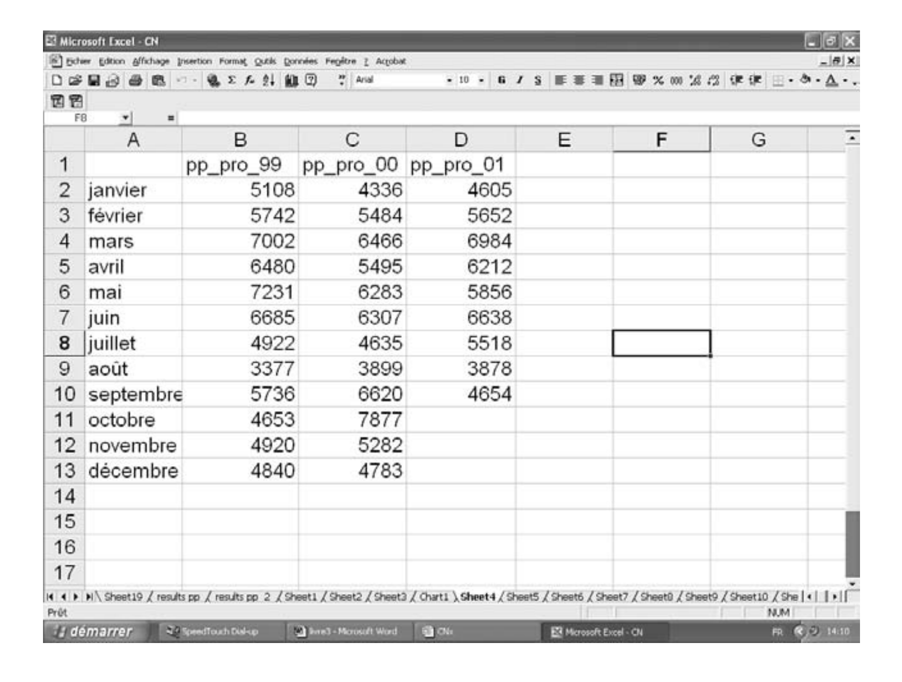
Les conclusions de l'examen graphique sont les suivantes :

1. L'évolution mensuelle des crédits de l'agence semble essentiellement due à des facteurs locaux plutôt que nationaux, ou à des facteurs spécifiques à l'entreprise et affectant sa part de marché dans la région.
2. L'évolution des crédits de l'agence semble présenter des fluctuations saisonnières.

Fluctuations saisonnières

Pour vérifier la présence de fluctuations saisonnières dans la série de production de crédits de l'agence bancaire, vous allez réaliser une nouvelle analyse graphique. Recopiez les données de la série pp_pro dans une nouvelle feuille du fichier BA, en la subdivisant en sous-variables correspondant à chaque année (voir figure 3.17).

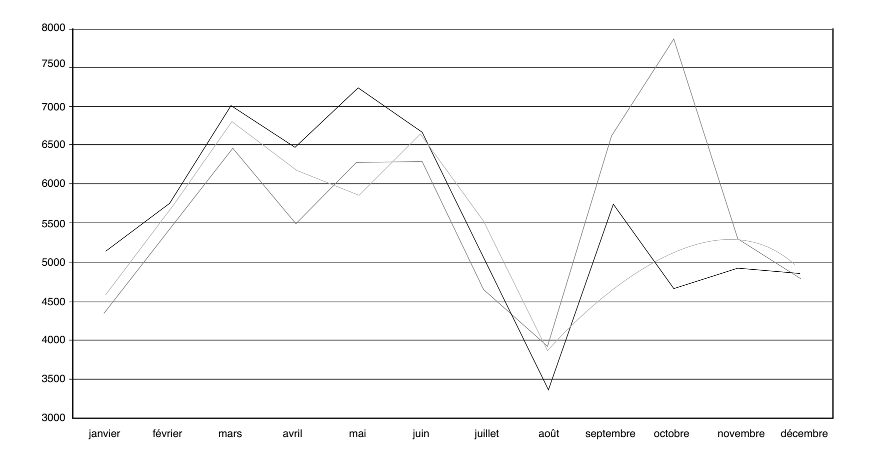
Figure 3.17



	A	B	C	D	E	F	G
1		pp_pro_99	pp_pro_00	pp_pro_01			
2	janvier	5108	4336	4605			
3	février	5742	5484	5652			
4	mars	7002	6466	6984			
5	avril	6480	5495	6212			
6	mai	7231	6283	5856			
7	juin	6685	6307	6638			
8	juillet	4922	4635	5518			
9	août	3377	3899	3878			
10	septembre	5736	6620	4654			
11	octobre	4653	7877				
12	novembre	4920	5282				
13	décembre	4840	4783				
14							
15							
16							
17							

Représentez ces sous-séries sous forme graphique en fonction de la première colonne, à savoir le mois (voir figure 3.18).

Figure 3.18



Un profil saisonnier systématique se répète clairement chaque année.

Modélisation et extrapolation

Sur la base des observations graphiques et d'un raisonnement économique élémentaire, spécifiez un modèle où l'évolution de pp_pro est expliquée par :

- des fluctuations saisonnières représentées par des variables dummy saisonnières ;
- le taux de marge, appliqué par la banque sur le coût de ses fonds et utilisé pour fixer le taux d'intérêt des crédits ;
- la valeur de pp_pro au mois précédent.

Le coefficient du taux de marge devrait être *a priori* négatif : si la banque réduit sa marge, et est donc prête à proposer des taux plus bas, elle peut attirer davantage d'emprunteurs. Vous devez d'abord préparer les données pour pouvoir estimer le modèle. À la suite des colonnes déjà remplies dans la feuille de données, recopiez en colonne K les dates à partir de décembre 1998, en commençant à la ligne 2. Puis, pour chaque date, complétez le tableau de la façon suivante :

- Recopiez la variable dépendante pp_pro en colonne L.
- Écrivez les onze variables dummy saisonnières, appelées $D1$, $D2$ . . . $D11$, en colonnes M à W.
- Recopiez la variable pp_pro_m en colonne X.
- Recopiez la variable dépendante retardée d'une période – appelez-la pp_pro(−1) – en colonne Y.

Les variables dummy saisonnières sont définies de la manière suivante :

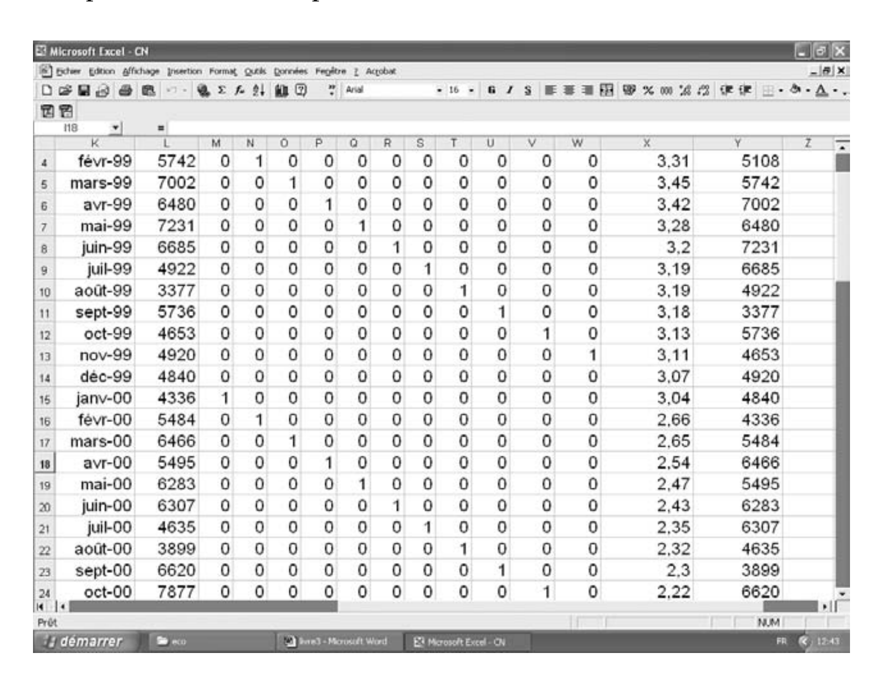
$D1$ vaut 1 pour janvier et 0 pour tous les autres mois.

$D2$ vaut 1 pour février et 0 pour tous les autres mois.

. . .

$D11$ vaut 1 pour novembre et 0 pour tous les autres mois.

Figure 3.19



	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
4	févr-99	5742	0	1	0	0	0	0	0	0	0	0	0	3,31	5108	
5	mars-99	7002	0	0	1	0	0	0	0	0	0	0	0	3,45	5742	
6	avr-99	6480	0	0	0	1	0	0	0	0	0	0	0	3,42	7002	
7	mai-99	7231	0	0	0	0	1	0	0	0	0	0	0	3,28	6480	
8	juin-99	6685	0	0	0	0	0	1	0	0	0	0	0	3,2	7231	
9	juil-99	4922	0	0	0	0	0	0	1	0	0	0	0	3,19	6685	
10	août-99	3377	0	0	0	0	0	0	0	1	0	0	0	3,19	4922	
11	sept-99	5736	0	0	0	0	0	0	0	0	1	0	0	3,18	3377	
12	oct-99	4653	0	0	0	0	0	0	0	0	0	1	0	3,13	5736	
13	nov-99	4920	0	0	0	0	0	0	0	0	0	0	1	3,11	4653	
14	déc-99	4840	0	0	0	0	0	0	0	0	0	0	0	3,07	4920	
15	janv-00	4336	1	0	0	0	0	0	0	0	0	0	0	3,04	4840	
16	févr-00	5484	0	1	0	0	0	0	0	0	0	0	0	2,66	4336	
17	mars-00	6466	0	0	1	0	0	0	0	0	0	0	0	2,65	5484	
18	avr-00	5495	0	0	0	1	0	0	0	0	0	0	0	2,54	6466	
19	mai-00	6283	0	0	0	0	1	0	0	0	0	0	0	2,47	5495	
20	juin-00	6307	0	0	0	0	0	1	0	0	0	0	0	2,43	6283	
21	juil-00	4635	0	0	0	0	0	0	1	0	0	0	0	2,35	6307	
22	août-00	3899	0	0	0	0	0	0	0	1	0	0	0	2,32	4635	
23	sept-00	6620	0	0	0	0	0	0	0	0	1	0	0	2,3	3899	
24	oct-00	7877	0	0	0	0	0	0	0	0	0	1	0	2,22	6620	

Commencer à la colonne K est un choix arbitraire. Vous pourriez commencer à la colonne I (voir figure 3.19).

Vous pouvez maintenant estimer le modèle. Cliquez successivement sur Outils (Tools)/Utilitaire d'analyse (Data analysis)/Régression linéaire (Regression) et remplissez

le tableau. Dans Plage pour la variable *Y*, écrivez $L$1:$L$35 pour définir les cellules des observations de la variable dépendante. Dans Plage pour les variables *X*, entrez $M$1:$Y$35 pour définir les cellules des observations des variables explicatives. Cochez Intitulé présent. Cochez Insérer une nouvelle feuille et écrivez results_pp. Cliquez sur OK. Les résultats s'affichent dans la feuille results_pp.

Puisque $R^2 = 0{,}765$, 76,5 % de la variabilité observée de pp_pro est expliquée par la variabilité des variables explicatives. Le coefficient estimé de la variable explicative vaut −137 ; il est donc négatif comme prévu : si la banque augmente ses marges, elle décourage des candidats emprunteurs. Toutefois, l'imprécision de l'estimation de ce coefficient, donnée par son écart type (standard errors) est énorme. Sa *t*-stat est donc petite et la probabilité critique (*p*-value) de sa *t*-stat est supérieure à 0,05 : vous ne pouvez rejeter l'hypothèse que le coefficient vrai de ll_pro_m soit nul. Pour autant, il n'est pas sûr qu'il soit nul : cela signifie simplement que 0 fait partie de l'ensemble de valeurs qui ne sont pas rejetées par le test (les valeurs d'un intervalle de confiance à 90 %). L'interprétation correcte du test est que le coefficient de ll_pro_m est compris dans un certain intervalle de confiance, avec une probabilité de 0,9. Toutefois, cet intervalle est trop large, si bien que vous ne pouvez montrer statistiquement si ce coefficient est négatif, nul ou positif.

Il peut alors être légitime de supprimer pp_pro_m des variables explicatives et d'estimer de nouveau le modèle. Supprimez donc la colonne X dans la feuille de données : la variable pp_pro(−1) vient occuper la colonne X. Cliquez sur Outils puis sur Régression linéaire. Dans Plage pour la variable *Y*, écrivez $L$1:$L$35 pour définir les cellules des observations de la variable dépendante. Dans Plage pour les variables *X*, écrivez $M$1:$X$35 pour définir les cellules des observations des variables explicatives. Cochez Intitulé présent puis Insérer une nouvelle feuille, et écrivez results_pp2. Cliquez sur OK. Les résultats s'affichent dans la feuille results_pp2.

Ensuite, utilisez ce modèle estimé pour prévoir la production de crédits prêts personnels d'octobre 2001 à décembre 2002. Dans la feuille de données, prolongez les dates à la colonne K et les variables dummy aux colonnes M à W, pour les lignes 36 à 50. Recopiez aussi les coefficients estimés de la feuille results_pp2 dans la feuille de données, aux cellules J36 à J48. Calculez alors la valeur de pp_pro en utilisant les coefficients estimés du modèle. En ce sens, dans la cellule L36, écrivez la formule pour pp_pro en octobre 2001 :

```
=$J$36+$J$37*M36+$J$38*N36+$J$39*O36+$J$40*P36+$J$41*Q36+$J$42*R36
+$J$43*S36+$J$44*T36+$J$45*U36+$J$46*V36+$J$47*W36+$J$48*L35
```

Dans la cellule L37, écrivez la formule pour pp_pro en novembre 2001 :

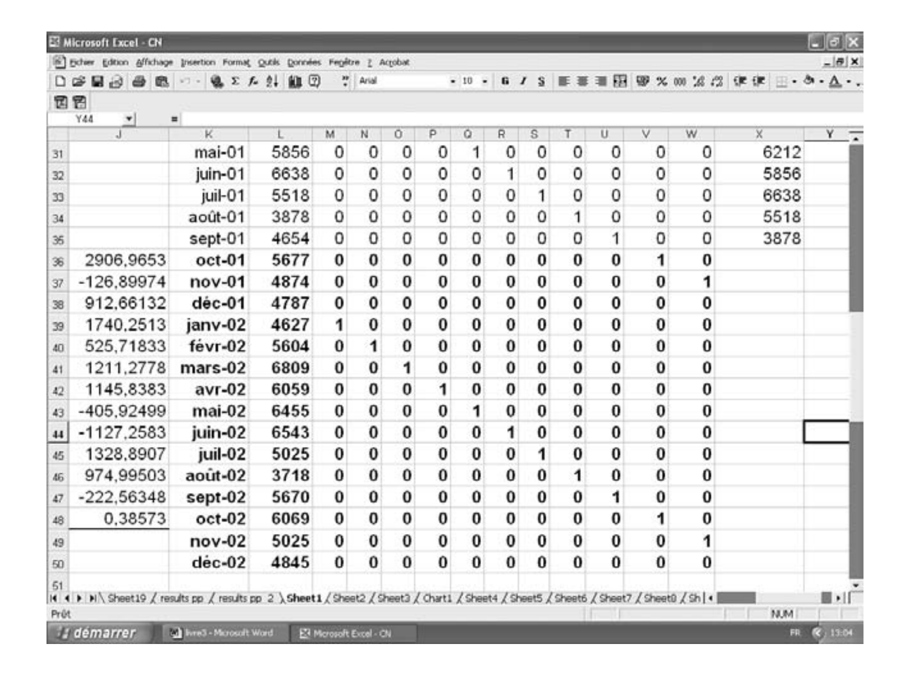
```
=$J$36+$J$37*M37+$J$38*N37+$J$39*O37+$J$40*P37+$J$41*Q37+$J$42*R37
+$J$43*S37+$J$44*T37+$J$45*U37+$J$46*V37+$J$47*W37+$J$48*L36
```

Réitérez jusqu'à la cellule L50, dans laquelle vous écrivez la formule pour pp_pro en décembre 2002 :

```
=$J$36+$J$37*M50+$J$38*N50+$J$39*O50+$J$40*P50+$J$41*Q50+$J$42*R50
 +$J$43*S50+$J$44*T50+$J$45*U50+$J$46*V50+$J$47*W50+$J$48*L49
```

Vous obtenez des valeurs extrapolées pour la production de prêts personnels (voir figure 3.20, page ci-contre).

Figure 3.20



	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
31		mai-01	5856	0	0	0	0	1	0	0	0	0	0	0	6212	
32		juin-01	6638	0	0	0	0	0	1	0	0	0	0	0	5856	
33		juil-01	5518	0	0	0	0	0	0	1	0	0	0	0	6638	
34		août-01	3878	0	0	0	0	0	0	0	1	0	0	0	5518	
35		sept-01	4654	0	0	0	0	0	0	0	0	1	0	0	3878	
36	2906,9653	oct-01	5677	0	0	0	0	0	0	0	0	0	1	0		
37	-126,89974	nov-01	4874	0	0	0	0	0	0	0	0	0	0	1		
38	912,66132	déc-01	4787	0	0	0	0	0	0	0	0	0	0	0		
39	1740,2513	janv-02	4627	1	0	0	0	0	0	0	0	0	0	0		
40	525,71833	févr-02	5604	0	1	0	0	0	0	0	0	0	0	0		
41	1211,2778	mars-02	6809	0	0	1	0	0	0	0	0	0	0	0		
42	1145,8383	avr-02	6059	0	0	0	1	0	0	0	0	0	0	0		
43	-405,92499	mai-02	6455	0	0	0	0	1	0	0	0	0	0	0		
44	-1127,2583	juin-02	6543	0	0	0	0	0	1	0	0	0	0	0		
45	1328,8907	juil-02	5025	0	0	0	0	0	0	1	0	0	0	0		
46	974,99503	août-02	3718	0	0	0	0	0	0	0	1	0	0	0		
47	-222,56348	sept-02	5670	0	0	0	0	0	0	0	0	1	0	0		
48	0,38573	oct-02	6069	0	0	0	0	0	0	0	0	0	1	0		
49		nov-02	5025	0	0	0	0	0	0	0	0	0	0	1		
50		déc-02	4845	0	0	0	0	0	0	0	0	0	0	0		
51																

EXERCICE 4 LE CAS PRODUCTEURS D'ÉLECTRICITÉ

Énoncé

Des investisseurs privés envisagent de financer une nouvelle usine de production d'électricité, d'une certaine capacité de production de kilowatts à l'heure. Avant de se décider et de fixer la taille de l'usine, ils souhaitent avoir la réponse à plusieurs questions, dont celles-ci : comment évoluent le coût total d'une centrale électrique en fonction de la quantité produite ? En d'autres termes, de quelle nature sont les rendements ? Quel est l'impact d'une augmentation de 1 % du prix du fuel sur le coût total, après une réorganisation optimale suite à cette augmentation ?

Comme informations, ils disposent en particulier de la production et du coût total de production des usines électriques déjà installées. Il s'agir des données de cent vingt-trois sociétés productrices d'électricité américaines rassemblées en 1970 et utilisées par L. Christensen et W. Greene [CHR 1984]. Elles se trouvent dans le fichier EL.xls téléchargeable sur le site Internet www.pearsoneducation.fr. Les variables sont les suivantes :

OUTPUT : la production de la centrale électrique ;
PRICELAB : le prix d'une unité de travail ;
PRICECAP : le prix d'une unité de capital physique ;
PFUEL : le prix d'une unité de fuel ;
COST : le coût total de production de la centrale.

Travaillez avec TSP.

Solution

Une production $Q = OUTPUT$ est fabriquée à partir de trois inputs : $x_1 = LABOR$, $x_2 = CAPITAL$, $x_3 = FUEL$, selon une technologie représentée par une certaine fonction de production : $Q = F(x_1, x_2, x_3)$. Supposez que la technologie de production d'électricité peut être représentée par une fonction de production Cobb-Douglas : $Q = Ax_1^{\gamma_1} x_2^{\gamma_2} x_3^{\gamma_3}$.

Les rendements d'échelle sont constants si $\gamma_1 + \gamma_2 + \gamma_3 = 1$. Ils sont croissants si $\gamma_1 + \gamma_2 + \gamma_3 > 1$ et décroissants si $\gamma_1 + \gamma_2 + \gamma_3 < 1$. En toute logique, vous pouvez supposer que les entreprises productrices d'électricité agissent rationnellement, et donc qu'elles minimisent leurs coûts de production. Cela veut dire qu'une entreprise, pour produire un montant Q d'électricité, choisit une quantité x_1 de travail, une quantité x_2 de capital et une quantité x_3 de fuel de manière à minimiser le coût de production. Chaque entreprise résout donc le problème : $\min_{x_1,x_2,x_3} p_1x_1 + p_2x_2 + p_3x_3$ sous la contrainte $Q = Ax_1^{\gamma_1}x_2^{\gamma_2}x_3^{\gamma_3}$, où p_1 est le prix d'une unité de travail, p_2 le prix d'une unité de capital et p_3 le prix d'une unité de fuel. Vous pouvez encore représenter ce problème avec un multiplicateur de Lagrange : $\min_{x_1,x_2,x_3,\lambda} p_1x_1 + p_2x_2 + p_3x_3 + \lambda\left(Q - Ax_1^{\gamma_1}x_2^{\gamma_2}x_3^{\gamma_3}\right)$. Les conditions de premier ordre de ce problème de choix optimal sont : $p_1 = \lambda\gamma_1 Ax_1^{\gamma_1-1}x_2^{\gamma_2}x_3^{\gamma_3}$, $p_2 = \lambda\beta_2 Ax_1^{\gamma_1}x_2^{\gamma_2-1}x_3^{\gamma_3}$, $p_3 = \lambda\beta_3 Ax_1^{\beta_1}x_2^{\gamma_2}x_3^{\gamma_3-1}$ et $Q = Ax_1^{\gamma_1}x_2^{\gamma_2}x_3^{\gamma_3}$. Ces conditions forment un système d'équations dont les solutions sont :

$$x_1 = \left(QA^{-1}\left(\frac{p_1}{p_2}\frac{\gamma_2}{\gamma_1}\right)^{-\gamma_2}\left(\frac{p_1}{p_3}\frac{\gamma_3}{\gamma_1}\right)^{-\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}$$

$$x_2 = \frac{p_1}{p_2}\frac{\gamma_2}{\gamma_1}\left(QA^{-1}\left(\frac{p_1}{p_2}\frac{\gamma_2}{\gamma_1}\right)^{-\gamma_2}\left(\frac{p_1}{p_3}\frac{\gamma_3}{\gamma_1}\right)^{-\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}$$

$$x_3 = \frac{p_1}{p_3}\frac{\gamma_3}{\gamma_1}\left(QA^{-1}\left(\frac{p_1}{p_2}\frac{\gamma_2}{\gamma_1}\right)^{-\gamma_2}\left(\frac{p_1}{p_3}\frac{\gamma_3}{\gamma_1}\right)^{-\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}$$

$$\begin{aligned}\lambda = {} & \frac{1}{\gamma_1+\gamma_2+\gamma_3}\,\frac{Q}{A}\left(p_1^{\gamma_1}p_2^{\gamma_2}p_3^{\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}\left(\frac{p_1}{p_3}\frac{\gamma_3}{\gamma_1}\right)^{-\gamma_3}\\ & \times\left(\left(\gamma_1^{\gamma_2+\gamma_3}\gamma_2^{-\gamma_2}\gamma_3^{-\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}} + \left(\gamma_1^{-\gamma_1}\gamma_2^{\gamma_1+\gamma_3}\gamma_3^{-\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}\right.\\ & \left. + \left(\gamma_1^{-\gamma_1}\gamma_2^{-\gamma_2}\gamma_3^{\gamma_1+\gamma_2}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}\right)\end{aligned}$$

Le coût total de production d'une quantité Q d'électricité, noté COST, est donné par :

$$COST \underset{\text{déf}}{=} p_1x_1 + p_2x_2 + p_3x_3$$

Cela implique que :

$$\begin{aligned}COST = {} & \frac{Q}{A}\left(p_1^{\gamma_1}p_2^{\gamma_2}p_3^{\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}\left(\frac{p_1}{p_3}\frac{\gamma_3}{\gamma_1}\right)^{-\gamma_3}\left(\left(\gamma_1^{\gamma_2+\gamma_3}\gamma_2^{-\gamma_2}\gamma_3^{-\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}\right.\\ & \left. + \left(\gamma_1^{-\gamma_1}\gamma_2^{\gamma_1+\gamma_3}\gamma_3^{-\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}} + \left(\gamma_1^{-\gamma_1}\gamma_2^{-\gamma_2}\gamma_3^{\gamma_1+\gamma_2}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}\right)\end{aligned}$$

Vous pouvez en déduire la relation en logarithmes suivante :

$$\ln COST = \beta_1 + \beta_2 \ln Q + \beta_3 \ln p_1 + \beta_4 \ln p_2 + \beta_5 \ln p_3$$

où

$$\begin{aligned}\beta_1 = \ln & \left(\left(\gamma_1^{\gamma_2+\gamma_3}\gamma_2^{-\gamma_2}\gamma_3^{-\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}} + \left(\gamma_1^{-\gamma_1}\gamma_2^{\gamma_1+\gamma_3}\gamma_3^{-\gamma_3}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}\right.\\ & \left. + \left(\gamma_1^{-\gamma_1}\gamma_2^{-\gamma_2}\gamma_3^{\gamma_1+\gamma_2}\right)^{\frac{1}{\gamma_1+\gamma_2+\gamma_3}}\right) - \frac{\ln A}{\gamma_1+\gamma_2+\gamma_3}\end{aligned}$$

$$\beta_2 = \frac{1}{\gamma_1+\gamma_2+\gamma_3}, \quad \beta_3 = \frac{\gamma_1}{\gamma_1+\gamma_2+\gamma_3},$$
$$\beta_4 = \frac{\gamma_2}{\gamma_1+\gamma_2+\gamma_3}, \quad \text{et} \quad \beta_5 = \frac{\gamma_3}{\gamma_1+\gamma_2+\gamma_3}$$

Si toutes les entreprises i ont la même technologie de production, la détermination de leurs coûts peut donc être représentée ainsi :

$$\ln COST_i = \beta_1 + \beta_2 \ln Q_i + \beta_3 \ln p_{1i} + \beta_4 \ln p_{2i} + \beta_5 \ln p_{3i} + u_i$$

pour tout $i = 1 \ldots 123$, où le terme d'erreur u_i représente la partie du logarithme du coût de l'entreprise i qui ne s'explique pas en fonction des logarithmes de la quantité produite et des prix des inputs, de la manière décrite par le raisonnement précédent. Il s'agit d'une régression linéaire sur des données en coupe instantanée. L'estimation des coefficients inconnus β_i, par la méthode des MCO, se fait au moyen des instructions TSP suivantes :

```
FREQ N;
SMPL 1 123;
READ(FILE='C:\EL.XLS');
LC=LOG(COST);
LQ=LOG(OUTPUT);
LL=LOG(PRICELAB);
LK=LOG(PRICECAP);
LP=LOG(PFUEL);
NOPLOT;
OLSQ LC C LQ LL LK LP;
```

Les résultats sont les suivants :

```
                         Equation 1
                        ============
          Method of estimation = Ordinary Least Squares

Dependent variable : LC
Current sample : 1 to 123
Number of observations : 123
       Mean of dep. var. = 2.96715      LM het. test = 1.93386 [.164]
  Std. dev. of dep. var. = 1.52498     Durbin-Watson = 1.33427 [<.000]
Sum of squared residuals = 5.35831  Jarque-Bera test = 63.3478 [.000]
   Variance of residuals = .045409   Ramsey's RESET2 = 150.297 [.000]
Std. error of regression = .213095   F (zero slopes) = 1532.51 [.000]
               R-squared = .981114    Schwarz B.I.C. = -6.15258
      Adjusted R-squared = .980474    Log likelihood = 18.1830
            Estimated     Standard
Variable  Coefficient     Error       t-statistic    P-value
C         -7.87891       1.29467       -6.08567       [.000]
LQ        .826268        .010949       75.4667        [.000]
LL        .099547        .130458       .763058        [.447]
LK        .192253        .130796       1.46987        [.144]
LP        .709348        .073380       9.66676        [.000]
```

Ignorez la valeur du test de Durbin et Watson puisque le concept d'autocorrélation des erreurs ne s'applique pas à des régressions utilisant des données en coupe instantanée. Les

99,11 % de la dispersion des logarithmes des coûts entre les différentes entreprises sont expliqués par la dispersion des logarithmes des variables explicatives (prix du travail, du capital et du fuel, production). Cette excellente performance du modèle le rend fiable pour la problématique considérée : vous pouvez l'utiliser pour répondre aux questions initiales. Par ailleurs, les rendements sont croissants dans l'industrie de production d'électricité puisque :

$$\hat{\gamma}_1 + \hat{\gamma}_2 + \hat{\gamma}_3 = \frac{1}{\hat{\beta}_2^{MCO}} = \frac{1}{0{,}826268} = 1{,}21026 > 1$$

Une augmentation de 1 % du prix du fuel entraîne une faible augmentation de 0,71 % du coût total de production, mais seulement après que la combinaison des quantités de facteurs ait été modifiée pour minimiser les coûts, en réponse à la nouvelle structure des prix.

EXERCICE 5 Le cas Prix des maisons

Énoncé

Vous disposez des données suivantes pour cinq cent quarante-six maisons mises en vente à Windsor, Canada :

PRIX : le prix de vente de la maison.

SUPERFICIE : la superficie de la parcelle sur laquelle se trouve la maison.

CHAMBRES : le nombre de chambres.

SDB : le nombre de salles de bain.

ETAGES : le nombre d'étages.

ALLEE : cette variable vaut 1 si la maison est dotée d'une allée privative, et 0 sinon.

SALLEJEU : cette variable vaut 1 si une salle de jeux est disponible, et 0 sinon.

GAZ : cette variable vaut 1 si un raccordement au gaz est disponible, et 0 sinon.

AIR : cette variable vaut 1 si un système d'air conditionné est installé, et 0 sinon.

GARAGES : le nombre de garages.

SITUATION : cette variable vaut 1 si la situation de la maison est particulièrement agréable, et 0 sinon.

Toutes ces variables se trouvent dans le fichier MAISON.xls téléchargeable à partir du site Internet www.pearsoneducation.fr. Ces données ont été utilisées par Paul Anglin et Ramazan Gencay [ANG 1996]. Mettez au point un outil permettant d'évaluer le prix de vente potentiel d'une maison, en fonction des valeurs de toutes les autres variables. Travaillez avec TSP.

Solution

Il faut d'abord spécifier un modèle linéaire reliant le prix de vente d'une maison à ses propriétés. Supposez que le prix de vente d'une maison est déterminé par le modèle suivant :

$$\begin{aligned} PRIX_i = {} & \beta_1 + \beta_2 SUPERFICIE_i + \beta_3 CHAMBRES_i + \beta_4 SDB_i + \beta_5 ETAGES_i + \beta_6 ALLEE_i \\ & + \beta_7 SALLEJEU_i + \beta_8 GAZ_i + \beta_9 AIR_i + \beta_{10} GARAGES_i + \beta_{11} SITUATION_i + u_i \end{aligned}$$

pour tout $i = 1 \ldots 546$. La variable u_i est un terme d'erreur inobservable.



Il s'agit d'un modèle linéaire reliant des variables en coupe instantanée. Il faut estimer les coefficients β_i par la méthode des MCO. Le programme d'instructions TSP est le suivant :

```
freq n;
smpl 1 546;
read(file='c :\maison.xls');
noplots;
olsq prix c superficie chambres sdb etages allee sallejeu gaz air garages situation;
```

Les résultats de l'estimation sont les suivants :

```
                              Equation 1
                             ============
                 Method of estimation = Ordinary Least Squares

Dependent variable : PRIX
Current sample : 1 to 546
Number of observations : 546
       Mean of dep. var. = 68121.6          LM het. test = 36.2854 [.000]
  Std. dev. of dep. var. = 26702.7         Durbin-Watson = 1.61923 [<.000]
Sum of squared residuals = .129829E+12  Jarque-Bera test = 263.929 [.000]
   Variance of residuals = .242672E+09   Ramsey's RESET2 = 20.5948 [.000]
Std. error of regression = 15577.9       F (zero slopes) = 106.635 [.000]
               R-squared = .665908        Schwarz B.I.C. = 6074.72
      Adjusted R-squared = .659663        Log likelihood = -6040.06
             Estimated     Standard
Variable    Coefficient      Error       t-statistic   P-value
C           -3127.96       3433.25       -.911079      [.363]
SUPERFICIE  3.45250        .352737       9.78776       [.000]
CHAMBRES    2341.89        1046.81       2.23716       [.026]
SDB         14819.3        1498.12       9.89194       [.000]
ETAGES      5674.82        897.823       6.32064       [.000]
ALLEE       6886.53        2064.94       3.33498       [.001]
SALLEJEU    6793.14        1797.79       3.77861       [.000]
GAZ         13016.0        3249.42       4.00564       [.000]
AIR         12855.3        1569.26       8.19195       [.000]
GARAGES     4287.96        848.882       5.05130       [.000]
SITUATION   10460.9        1654.98       6.32085       [.000]
```

Ce modèle peut être utilisé par des agents immobiliers lorsqu'ils doivent fixer le prix de vente de départ d'une habitation. Il faut d'abord rassembler les valeurs des différentes variables de droite (superficie...) pour la maison à vendre. Il suffit ensuite d'introduire les valeurs de ces variables dans la formule suivante :

$$\begin{aligned}\text{Prix de vente de départ} = {} & \beta_1 + \beta_2 SUPERFICIE_i + \beta_3 CHAMBRES_i + \beta_4 SDB_i \\ & + \beta_5 ETAGES_i + \beta_6 ALLEE_i + \beta_7 SALLEJEU_i \\ & + \beta_8 GAZ_i + \beta_9 AIR_i + \beta_{10} GARAGES_i \\ & + \beta_{11} SITUATION_i\end{aligned}$$

Remplacez les coefficients inconnus β_i par leurs valeurs estimées du tableau de résultats.

Exercices

EXERCICE 6 LE CAS PRIX DES HÔTELS

Énoncé

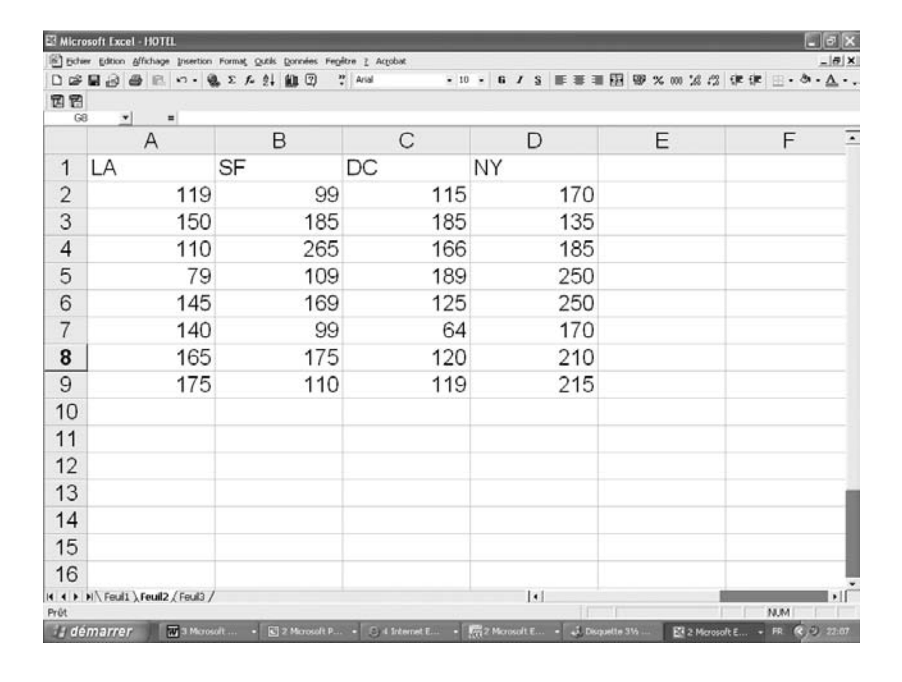
Vous souhaitez savoir si le prix d'une nuitée dans une chambre d'hôtel d'une grande ville américaine dépend de la ville dans laquelle l'hôtel est situé, à qualité égale. Pour répondre à cette question, vous allez analyser des données des villes Los Angeles, Sans Francisco, Washington DC et New York. Pour chacune d'elles, vous disposez du prix d'une nuitée et du nombre d'étoiles de huit hôtels choisis au hasard dans cette ville. Les données se trouvent initialement dans la feuille 1 du classeur Excel Hotel.xls. Ce fichier est téléchargeable sur le site www.pearsoneducation.fr. Les colonnes B, A et C contiennent respectivement les en-têtes Hotel, City, et Price en première ligne, puis les noms des différents hôtels, la ville où ils se situent et le prix d'une nuitée qu'ils tarifent sur les lignes suivantes.

Solution

Solution avec un tableur

Il s'agit d'une application de l'analyse de variance, ou ANOVA ; elle n'est qu'un cas particulier de la méthode des *MCO* avec variables indicatrices (ou dummy). Cette ANOVA est à un facteur. Pour pouvoir utiliser la fonction automatique Analyse de variance : un facteur d'Excel, recopiez d'abord les données nécessaires dans une autre feuille du classeur, chaque modalité de la caractéristique occupant une colonne (voir figure 3.21).

Figure 3.21



	A	B	C	D	E	F
1	LA	SF	DC	NY		
2	119	99	115	170		
3	150	185	185	135		
4	110	265	166	185		
5	79	109	189	250		
6	145	169	125	250		
7	140	99	64	170		
8	165	175	120	210		
9	175	110	119	215		
10						
11						
12						
13						
14						
15						
16						

Sur cette nouvelle feuille appelée Feuil2, cliquez sur Outils puis sur Utilitaire d'analyse. Dans le menu déroulant, sélectionnez Analyse de variance : 1 facteur et cliquez sur OK. Cochez Intitulé en première ligne parce que les modalités (intitulés) de la caractéristique qualitative sont bien en première ligne. Écrivez A1$:$D$9 dans Plage d'entrée puisque ce sont les cellules contenant les données. Saisissez 0,05 comme Seuil de signification. Cochez Colonnes puis Insérer une nouvelle feuille, et nommez-la Anova 1 facteur. Cliquez sur OK. Les résultats s'affichent dans cette nouvelle feuille. La statistique *F* vaut 3,60702982.

La probabilité critique est 0,02548955, inférieure à 0,05. Rejetez donc l'hypothèse que l'espérance du prix d'une nuitée est identique pour les quatre villes. Ce prix dépend de la ville d'accueil.

Solution avec TSP

L'ANOVA à un facteur est une simple application de la régression linéaire. Sachant que :

- La variable dépendante est le vecteur des prix observés pour les trente-deux hôtels.
- Il n'y a pas de constante.
- Les variables explicatives sont quatre variables dummy se rapportant chacune à une ville.
- La variable dummy d'une ville particulière a trente-deux observations qui valent 0 si l'hôtel n'est pas dans cette ville et 1 si l'hôtel est dans cette ville.

... estimez par *MCO* le modèle suivant :

$$Prix_i = \beta_1 DU1_i + \beta_2 DU2_i + \beta_3 DU3_i + \beta_4 DU4_i + u_i\,, \quad \text{pour} \;\; i = 1 \ldots 32$$

... et testez l'hypothèse $\beta_1 = \beta_2 = \beta_3 = \beta_4$ par un test F habituel. Ce test n'est valable que si u est distribué normalement. Chaque coefficient β_j représente l'espérance de prix, $E(Prix_i)$, pour les hôtels de la ville j.

Pour estimer ce modèle avec TSP, suivez la procédure décrite au chapitre 2. Recopiez la colonne C du fichier Hotel.xls dans la colonne A d'un nouveau fichier que vous enregistrerez dans C:\ comme une feuille de calcul Excel, sous le nom de HO. Ensuite, utilisez les instructions suivantes sous TSP :

```
FREQ N;
SMPL 1 32;
READ(FILE='C :\HO.XLS');
LOAD DU1;
1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
LOAD DU2;
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
LOAD DU3;
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0;
LOAD DU4;
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1;
NOPLOT;
OLSQ PRICE DU1 DU2 DU3 DU4;
```

Ce programme produit les résultats suivants :

```
Dependent variable: PRICE
Current sample: 1 to 32
Number of observations: 32
       Mean of dep. var. = 155.063      LM het. test = .390756E-03 [.984]
  Std. dev. of dep. var. = 49.4681     Durbin-Watson = 1.95603 [<.643]
Sum of squared residuals = 54714.5  Jarque-Bera test = .833548 [.659]
   Variance of residuals = 1954.09   Ramsey's RESET2 = .850706E+38 [.000]
Std. error of regression = 44.2051   F (zero slopes) = 3.60703 [.025]
```

```
              R-squared = .278743    Schwarz B.I.C. = 171.444
     Adjusted R-squared = .201465    Log likelihood = -164.512
            Estimated    Standard
Variable   Coefficient    Error     t-statistic    P-value
DU1        135.375       15.6289     8.66186        [.000]
DU2        151.375       15.6289     9.68561        [.000]
DU3        135.375       15.6289     8.66186        [.000]
DU4        198.125       15.6289     12.6769        [.000]
```

Vous retrouvez dans ces résultats la statistique F avec sa probabilité critique.

Exercice 7 Consommation et simultanéité

Énoncé

Comme au chapitre 1, il faut estimer un modèle linéaire expliquant le taux de croissance de la consommation réelle en fonction d'une constante, du taux de croissance du revenu disponible réel et du taux d'inflation, en utilisant les données du fichier USA.XLS. Tenez compte d'une probable simultanéité entre consommation et revenu, qui pourrait entraîner une dépendance entre le terme d'erreur du modèle et une variable explicative (le taux de croissance du revenu). Il faut donc estimer le modèle en utilisant des variables instrumentales, qui représentent la constante, les taux de croissance retardés de la consommation, du revenu et des prix, ainsi que le taux de chômage UR. Réalisez cette application avec TSP.

Solution

Le programme permettant de réaliser cette estimation avec TSP est le suivant :

```
FREQ A;
SMPL 1960 1994;
READ(FILE='C :\USA.XLS');
SMPL 1961 1994;
DLC=LOG(CT)-LOG(CT(-1));
DLY=LOG(Y)-LOG(Y(-1));
DLP=LOG(P)-LOG(P(-1));
REGOPT(PVPRINT,STARS,LMLAGS=2,QLAGS=2) ALL;
NOPLOT;
SMPL 1962 1994;
INST DLC C DLY DLP INVR C DLY(-1) DLP(-1) DLC(-1) UR;
OLSQ DLC C DLY DLP;
```

Certaines variables instrumentales choisies étant des variables en taux de croissance retardés d'une période, l'estimation par VI doit se faire à partir de 1962 plutôt que 1961. Pour pouvoir comparer valablement les résultats à ceux obtenus par variables instrumentales, vous devez donc réaliser l'estimation par *MCO* à partir de 1962.

Les résultats de l'estimation par variables instrumentales sont :

```
                              Equation 1
                             ============
              Method of estimation = Instrumental Variable
Dependent variable: DLC
```

```
Endogenous variables: DLY DLP
Included exogenous variables: C
Excluded exogenous variables: DLY(-1) DLP(-1) DLC(-1) UR
Current sample: 1962 to 1994
Number of observations: 33
       Mean of dep. var. = .032166
  Std. dev. of dep. var. = .016776
Sum of squared residuals = .279486E-02
   Variance of residuals = .931620E-04
Std. error of regression = .965205E-02
               R-squared = .696993
      Adjusted R-squared = .676793
           Durbin-Watson = 1.77465 [.066,.565]
                  E'PZ*E = .256476E-03
           Estimated    Standard
Variable  Coefficient    Error      t-statistic     P-value
C          .023945     .011354      2.10896       * [.035]
DLY        .561161     .224709      2.49728       * [.013]
DLP        -.761937    .170785      -4.46138     ** [.000]
```

Comparez ces résultats avec ceux obtenus par la méthode des moindres carrés ordinaires :

```
                               Equation 2
                              ============
               Method of estimation = Ordinary Least Squares

Dependent variable: DLC
Current sample: 1962 to 1994
Number of observations: 33
          Mean of dep. var. = .032166
    Std.  dev. of dep. var. = .016776
   Sum of squared residuals = .269899E-02
      Variance of residuals = .899663E-04
   Std. error of regression = .948506E-02
                  R-squared = .700307
         Adjusted R-squared = .680327
               LM het. test = .401618 [.526]
              Durbin-Watson = 1.98694 [.334,.638]
Breusch/Godfrey LM : AR/MA1 = .135113E-02 [.971]
Breusch/Godfrey LM : AR/MA2 = .064560 [.968]
     Ljung-Box Q-statistic1 = .121645E-02 [.972]
     Ljung-Box Q-statistic2 = .089390 [.956]
  Wald nonlin. AR1 vs. lags = 1.36760 [.505]
                 ARCH test = .421448 [.516]
                CuSum test = .423037 [.853]
              CuSumSq test = .086548 [1.00]
                 Chow test = .513721 [.676]
     LR het. test (w/ Chow) = -1.40845 [1.00]
            White het. test = 5.31376 [.379]
```

```
             Jarque-Bera test = .515258 [.773]
            Shapiro-Wilk test = .965642 [.370]
              Ramsey's RESET2 = 4.49155 * [.043]
             F (zero slopes) = 35.0512 ** [.000]
              Schwarz B.I.C. = -103.218
    Akaike Information Crit. = -105.463
              Log likelihood = 108.463
            Estimated     Standard
Variable    Coefficient    Error         t-statistic    P-value
C          .019425        .652456E-02    2.97724     ** [.006]
DLY        .671728        .110019        6.10557     ** [.000]
DLP        -.849897       .101655        -8.36062    ** [.000]
```

Les coefficients estimés par la méthode des variables instrumentales sont présentés dans la colonne estimated coefficient du tableau de résultats ayant comme titre equation 1. Les coefficients estimés par la méthode des *MCO* sont présentés dans la colonne estimated coefficient du tableau de résultats ayant comme titre equation 2. Les coefficients estimés par VI et par *MCO* sont différents (même si cette différence est d'une ampleur modérée), ce qui suggère un éventuel problème de simultanéité.

EXERCICE 8 CONSOMMATION PAR LA MÉTHODE DU MAXIMUM DE VRAISEMBLANCE

Énoncé

Utilisez de nouveau les données du fichier USA.xls, téléchargeable à partir du site Internet afférent à ce livre. Estimez le modèle linéaire expliquant la variation logarithmique de la consommation en fonction des variations logarithmiques des revenus et des prix, mais cette fois en utilisant la méthode du maximum de vraisemblance. Supposez que le terme d'erreur est indépendant des variables de droite, qu'il s'agit d'un bruit blanc et qu'il est normal.

Solution

Le modèle à estimer est $DLC_t = \beta_1 + \beta_2 DLY_t + \beta_3 DLP_t + u_t$, où $u_t N(0, \sigma_u^2)$. Les u_t sont supposés être des bruits blancs et indépendants des DLC_t et DLY_t. Le logarithme de la fonction de vraisemblance est donc égal à :

$$\ln L\left(\beta, \sigma_u^2\right) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma_u^2 - \sum_{t=1}^{n} \frac{(Y_t - \beta_1 - \beta_2 DLY_t - \beta_3 DLP_t)^2}{2\sigma_u^2}$$

Il faut écrire un programme d'instructions TSP qui trouve les valeurs des coefficients qui maximisent cette fonction. Bien entendu, les valeurs qui maximisent $\ln\left(L(\beta, \sigma_u^2)\right)$ sont les mêmes que celles qui maximisent sa transformation monotone croissante suivante :

$$2\ln L\left(\beta, \sigma_u^2\right) + n\ln 2\pi = -n\ln \sigma_u^2 - \sum_{t=1}^{n} \frac{(Y_t - \beta_1 - \beta_2 DLY_t - \beta_3 DLP_t)^2}{\sigma_u^2}$$



Le programme d'instructions TSP est le suivant :

```
FREQ A;
SMPL 60 94;
READ(FILE='c :\USA.XLS');
LC=LOG(CT);
LY=LOG(Y);
LP=LOG(P);
SMPL 61 94;
DLC=LC-LC(-1);
DLY=LY-LY(-1);
DLP=LP-LP(-1);
FRML EQ1 LOGL=LOG(SIGINV)+LNORM((DLC--XB)*SIGINV);
FRML EQXB1 XB=B1+B2*DLY+B3*DLP;
EQSUB(NAME=OLS) EQ1 EQXB1;
PARAM B1 B2 B3 SIGINV;
SET SIGINV=1;
ML(HITER=N,HCOV=NBW) OLS;
```

Les coefficients à estimer sont `B1`, `B2` et `B3` qui représentent β_1, β_2 et β_3. Il faut également estimer σ_u. Cela revient à estimer σ_u^{-1}, représenté ici par `SIGINV`. `LOGL` représente la partie du logarithme de la fonction de vraisemblance qui correspond à une observation. La fonction `LNORM(x)` représente le logarithme de la fonction de densité d'une $N(0, 1)$ évaluée en x. Vous obtenez les résultats suivants :

```
                    MAXIMUM LIKELIHOOD ESTIMATION
                    =============================

EQUATION: OLS
Working space used: 645
                           STARTING VALUES
                    B1            B2            B3         SIGINV
VALUE          0.00000       0.00000       0.00000       1.00000
F=    31.266    FNEW=   7.9344    ISQZ=  0 STEP=  1.0000    CRIT=  34.236
F=    7.9344    FNEW=  -10.607    ISQZ=  0 STEP=  1.0000    CRIT=  38.141
F=  -10.607     FNEW=  -11.461    ISQZ=  3 STEP= 0.12500    CRIT=  378.75
F=  -11.461     FNEW=  -48.508    ISQZ=  1 STEP= 0.50000    CRIT=  117.74
F=  -48.508     FNEW=  -64.420    ISQZ=  1 STEP= 0.50000    CRIT=  52.250
F=  -64.420     FNEW=  -80.388    ISQZ=  2 STEP= 0.25000    CRIT=  87.521
F=  -80.388     FNEW=  -90.705    ISQZ=  0 STEP=  1.0000    CRIT=  34.624
F=  -90.705     FNEW=  -105.48    ISQZ=  2 STEP= 0.25000    CRIT=  126.22
F=  -105.48     FNEW=  -110.49    ISQZ=  1 STEP= 0.50000    CRIT=  19.698
F=  -110.49     FNEW=  -110.74    ISQZ=  0 STEP=  1.0000    CRIT= 0.50034
F=  -110.74     FNEW=  -110.75    ISQZ=  0 STEP=  1.0000    CRIT= 0.22757E-02
F=  -110.75     FNEW=  -110.75    ISQZ=  0 STEP=  1.0000    CRIT= 0.80716E-07
CONVERGENCE ACHIEVED AFTER 12 ITERATIONS

  34 FUNCTION EVALUATIONS.
Number of observations = 34        Log likelihood = 110.746
        Schwarz B.I.C. = -103.693
                          Standard
Parameter   Estimate       Error        t-statistic    P-value
B1          .016282      .613489E-02    2.65396       [.008]
```

Exercices

```
B2          .696019     .107093       6.49918       [.000]
B3          -.832804    .099321      -8.38500       [.000]
SIGINV      107.359     13.0191      8.24621        [.000]
Standard Errors computed from analytic second derivatives
(Newton)
                        Standard
Parameter   Estimate      Error       t-statistic   P-value
B1          .016282     .620506E-02   2.62395       [.009]
B2          .696019     .110242       6.31354       [.000]
B3          -.832804    .123473       -6.74481      [.000]
SIGINV      107.359     16.6155       6.46136       [.000]
Standard Errors computed from covariance of analytic first derivatives
(BHHH)
                        Standard
Parameter   Estimate      Error       t-statistic   P-value
B1          .016282     .641869E-02   2.53662       [.011]
B2          .696019     .108838       6.39501       [.000]
B3          -.832804    .082664       -10.0746      [.000]
SIGINV      107.359     11.1076       9.66537       [.000]
Standard Errors computed from analytic first and second derivatives
(Eicker-White)
```

EXERCICE 9 MODÉLISATION DE LA POLITIQUE MONÉTAIRE

Énoncé

Estimez un modèle expliquant le niveau des taux d'intérêts américains en fonction du taux d'inflation et du taux de chômage. Pour que les écarts types estimés des coefficients estimés soient corrects, même en présence d'hétéroscédasticité, utilisez une estimation robuste de la matrice de variance et de covariance des coefficients estimés. Comparez avec les écarts types obtenus par l'estimation habituelle de cette matrice.

Le fichier USA.xls téléchargeable sur le site www.pearsoneducation.fr rassemble plusieurs variables, de 1960 à 1994, dont :

R : le taux d'intérêt ;

UR : le taux de chômage ;

P : l'indice des prix à la consommation.

Travaillez avec TSP.

Solution

Vous allez estimer un modèle linéaire dont la variable dépendante est le taux d'intérêt R_t, et les variables explicatives sont l'inflation INF_t, le taux de chômage UR_t et le taux d'intérêt retardé R_{t-1}. Le taux d'inflation INF_t doit être généré à partir de l'indice des prix P, à l'aide de la formule $INF_t = 100(P_t - P_{t-1})/P_{t-1}$. Le programme permettant d'estimer cette relation par moindres carrés ordinaires est présenté ci-après. Le modèle est estimé successivement sans et avec une matrice de variance et de covariance robuste.

```
freq a;
smpl 1960 1994;
read(file='c :\usa.xls');
smpl 1961 1994;
inf=100*(p-p(-1))/p(-1);
regopt(pvprint,stars) all;
```



```
supres vcov vcor csmax csqmax chow;
noplot;
olsq r c inf ur r(-1);
olsq(robustse, hctype=1) r c inf ur r(-1);
```

L'option `robustse` commande une estimation robuste des écarts types. L'option `hctype=` permet de choisir entre différentes formules. Ici, vous utilisez une formule dont les degrés de liberté sont ajustés pour un petit échantillon. L'instruction `supres` et ses arguments suppriment certains résultats, dont l'impression des tests de constance des coefficients, qui sont inutiles.

```
                          Equation 1
                         ============
              Method of estimation = Ordinary Least Squares

Dependent variable: R
Current sample: 1961 to 1994
Number of observations: 34
        Mean of dep. var. = 7.64232
  Std. dev.  of dep. var. = 2.61989
 Sum of squared residuals = 23.2655
    Variance of residuals = .775517
 Std. error of regression = .880634
                R-squared = .897285
       Adjusted R-squared = .887014
             LM het. test = 4.81378 * [.028]
            Durbin-Watson = 1.97024 [.247,.694]
               Durbin's h = -.298590 [.765]
           Durbin's h alt. = -.291388 [.771]
                ARCH test = .644711E-02 [.936]
   LR het. test (w/ Chow) = 18.7015 ** [.000]
          White het. test = 12.1299 [.206]
         Jarque-Bera test = 4.35856 [.113]
        Shapiro-Wilk test = .936582 * [.049]
          Ramsey's RESET2 = 2.89648 [.099]
          F (zero slopes) = 87.3571 ** [.000]
          Schwarz B.I.C.  = 48.8470
Akaike Information Crit.  = 45.7943
           Log likelihood = -41.7943
           Estimated     Standard
Variable  Coefficient      Error       t-statistic     P-value
C          .916789       .655369       1.39889         [.172]
INF        .279208       .069364       4.02527      ** [.000]
UR        -.159064       .151533      -1.04969         [.302]
R(-1)      .846701       .091481       9.25547      ** [.000]
```

```
                    Equation 2
                   ============
           Method of estimation = Ordinary Least Squares
            Estimated     Standard
Variable   Coefficient      Error      t-statistic     P-value
C          .916789        .384744      2.38286       * [.024]
INF        .279208        .082467      3.38568      ** [.002]
UR         -.159064       .083728      -1.89976        [.067]
R(-1)      .846701        .098165      8.62526      ** [.000]
Standard Errors are heteroskedastic-consistent (HCTYPE=1).
```

Très logiquement, le coefficient de INF est positif tandis que celui de UR est négatif. Plusieurs tests d'hétéroscédasticité ont des probabilités critiques inférieures à 0,05, ce qui suggère une hétéroscédasticité du terme d'erreur. Cela implique que les écarts types estimés de manière conventionnelle, présentés sous le titre equation 1, ne sont pas fiables. Par contre, les écarts types estimés de manière robuste, présentés sous le titre equation 2, sont fiables. Par ailleurs, la précision de l'estimation du coefficient du taux de chômage semble bien meilleure quand les écarts types sont estimés de manière robuste.



Références bibliographiques

[ANG 1996] P. Anglin, R Gencay, Semiparametric Estimation of a Hedonic Price Function, *Journal of Applied Econometrics*, 11 (6), p. 633–648, 1996.

[BRE 1981] T.S. Breusch, L.G. Godfrey, A Review of Recent Work on Testing for Autocorrelation in Dynamic Simultaneous Models, dans *Macroeconomic Analysis : Essays in Macroeconomics and Econometrics*, eds D. Currie, R. Nobay and D. Peel, Croom Helm, London, 1981.

[BRE 1980] T.S. Breusch, A.R. Pagan, The Lagrange Multiplier Test and its Application to Model Specifications in Econometrics, *Review of Economic Studies*, 47, p. 239–253, 1980.

[CHR 1984] L. Christensen, W. Greene, Economies of Scale in US Electric Power Generation, *Journal of Political Economy*, 84 (4), p. 655–676, 1984.

[DAV 1993] R. Davidson, J.G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993.

[GRE 2000] W.H. Greene, *Econometric Analysis*, Fourth Edition, Prentice Hall International, New Jersey, 2000.

[MAC 1985] J.G. MacKinnon, H. White, Some Heteroskedasticity-Consistent Matrix Estimators with Improved Finite Sample Properties, *Journal of Econometrics*, 29, p. 305–325, 1985.

[NEW 1987] W.K. Newey, K.D. West, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, p. 703–708, 1987.

[PES 1999] M.H. Pesaran, L.W. Taylor, Diagnostics for IV Regressions, *Oxford Bulletin of Economics and Statistics*, 61, p. 255–281, 1999.

[PES 1994] M.H. Pesaran, R.J. Smith, A Generalized R Criterion for Regression Models Estimated by the Instrumental Variables Method, *Econometrica*, 62, p. 705–710, 1994.

[SPA 1986] A. Spanos, *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge, 1986.

[WHI 1980] H. White, A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, p. 817–838, 1980.



Équations multiples en univers stationnaire

Équations multiples en univers stationnaire

Ce chapitre traite des modèles économétriques à plusieurs équations. Il s'intéresse d'abord aux régressions simultanées et à la méthode SUR, puis aborde les modèles à équations simultanées, définit les notions importantes de forme structurelle et de forme réduite, avant de décrire les conditions d'identification des paramètres. Il passe ensuite en revue quelques méthodes d'estimation des modèles à équations simultanées : les doubles moindres carrés, les triples moindres carrés et le maximum de vraisemblance à information complète.

1 Régressions simultanées et méthode SUR

L'objectif de cette section est d'étudier l'estimation d'un modèle à plusieurs équations, ayant chacune des variables explicatives différentes, et dont les termes d'erreur contemporains sont corrélés entre eux, mais indépendants des variables explicatives. Un tel modèle est composé de m équations de type :

$$Y_i = X_i\beta_i + u_i \qquad (4.1)$$

où $Y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in} \end{pmatrix}$, $\beta_i = \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{ik_i} \end{pmatrix}$, $u_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{in} \end{pmatrix}$ et $X_i = \begin{pmatrix} 1 & X_{i21} & \cdots & X_{ik_i1} \\ 1 & X_{i22} & \cdots & X_{ik_i2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{i2n} & \cdots & X_{ik_in} \end{pmatrix}$ pour tout $i = 1 \ldots m$.

Les termes d'erreur contemporains des différentes équations sont corrélés entre eux : $\text{Cov}(u_{it}, u_{jt}) = \sigma_{ij}$ pour tout $i = 1 \ldots m$ et $j = 1 \ldots m$. En revanche, pris à des périodes différentes, ils ne sont pas corrélés entre eux : $\text{Cov}(u_{it}, u_{jt-\theta}) = 0$ pour tout $i = 1 \ldots m$, $j = 1 \ldots m$ et pour tout $\theta \geqslant 1$.

Le terme d'erreur de chaque équation est supposé homoscédastique et non autocorrélé : $\text{Cov}(u_{it}, u_{it-\theta}) = 0$ et $V(u_{it}) = \sigma_i^2$ pour tout$i = 1 \ldots m$. Ces hypothèses impliquent que :

$$\Sigma_{u_iu_j} = E(u_iu_j') = \begin{pmatrix} \sigma_{ij} & 0 & \cdots & 0 \\ 0 & \sigma_{ij} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{ij} \end{pmatrix} = \sigma_{ij} I_n$$

et

$$\Sigma_{u_i} = E(u_iu_i') = \begin{pmatrix} \sigma_i^2 & 0 & \cdots & 0 \\ 0 & \sigma_i^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_i^2 \end{pmatrix} = \sigma_i^2 I_n \qquad (4.2)$$

Ces équations sont appelées « équations SUR » (Seemingly Unrelated Regression) par Zellner [ZEL 1962]. La méthode SUR permet d'obtenir des estimations plus précises des coefficients que celles obtenues en appliquant simplement les *MCO* à chaque équation. En « raccordant bout à bout » les vecteurs Y_i des m équations, on obtient une variable dépendante composite ayant $m \times n$ lignes et 1 colonne :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = (Y_{11}Y_{12} \ldots Y_{1n}Y_{21}Y_{22} \ldots Y_{2n} \ldots Y_{m1}Y_{m2} \cdots Y_{mn})'$$

Le système d'équations initial peut s'écrire sous la forme d'une seule équation linéaire :

$$Y = X\beta + u \qquad (4.3)$$

où $X = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_m \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}$ et $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix}$.

La matrice X a $m \times n$ lignes et $k_1 + k_2 + \cdots + k_m$ colonnes. Le vecteur β a $k_1 + k_2 + \cdots + k_m$ lignes et 1 colonne. Le vecteur u a $m \times n$ lignes et 1 colonne. La matrice de variance et

covariance du terme d'erreur u est :

$$\Sigma_u = E(uu_i') = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_m^2 \end{pmatrix} \otimes I_n = \Sigma \otimes I_n \tag{4.4}$$

Dans l'équation (4.4), on utilise le produit de Kronecker (ou produit tensoriel). Si les valeurs des σ_i^2 et σ_{ij} sont connues, il est logique d'estimer ce modèle par la méthode des moindres carrés généralisés, dite MCG (voir chapitre 3). Toutefois, elles ne sont pas connues dans la réalité. Zellner [ZEL 1962] propose d'estimer d'abord chaque équation séparément par MCO :

$$Y_i = X_i\hat{\beta}_i^{MCO} + e_i \tag{4.5}$$

et d'utiliser les résidus de ces équations pour obtenir des estimations des σ_i^2 et σ_{ij} :

$$\hat{\sigma}_i^2 = \frac{e_i'e_i}{n-k_i} \quad \text{et} \quad \hat{\sigma}_{ij} = \frac{e_i'e_j}{\sqrt{n-k_i}\sqrt{n-k_j}} \tag{4.6}$$

Ces estimations forment une matrice $\hat{\Sigma}$ estimée, utilisable dans la formule des MCG. On obtient ainsi les estimateurs SUR des coefficients :

$$\hat{\beta}^{SUR} = \left(X'\left(\hat{\Sigma}^{-1} \otimes I_n\right)X\right)^{-1} X'\left(\hat{\Sigma}^{-1} \otimes I_n\right)Y \tag{4.7}$$

On montre que ces estimateurs sont plus précis (variance plus faible) que ceux que l'on obtient en estimant chaque équation indépendamment par *MCO*. L'explication est simple : la méthode *SUR* tient compte des corrélations contemporaines entre les termes d'erreur des différentes équations. Toutefois, si les covariances contemporaines sont toutes nulles, ou si les variables explicatives sont les mêmes dans les m équations ($X_1 = X_2 = \cdots = X_m$), les estimateurs *SUR* et les estimateurs de *MCO* de chaque équation prise indépendamment sont égaux.

2 Systèmes d'équations simultanées et identification

Pour comprendre ce qu'est un modèle à équations simultanées, il faut distinguer les variables **endogènes** des variables **exogènes**. Les premières sont telles que leur valeur en t est déterminée par le modèle alors que les secondes ne sont pas déterminées par le modèle : elles sont données telles quelles et, conditionnellement à leur valeur, le modèle détermine les valeurs des variables endogènes. Les variables exogènes et endogènes retardées sont dites **prédéterminées**.

Un modèle à équations simultanées décrit la manière dont se déterminent simultanément les valeurs de G variables endogènes à la période t, en fonction des valeurs contemporaines ou retardées de certaines variables exogènes, et d'éventuelles valeurs retardées des variables endogènes. Certaines variables endogènes sont fonction d'autres variables endogènes et cette simultanéité produit des **équations apparemment linéaires**, où le terme d'erreur est lié aux variables explicatives.

Un modèle à équations simultanées est décrit par G **équations structurelles** de la façon suivante :

$$\beta_{i1}Y_{1t} + \beta_{i2}Y_{2t} + \cdots + \beta_{i3}Y_{Gt} + \gamma_{i1}X_{1t} + \gamma_{i2}X_{2t} + \cdots + \gamma_{iK}X_{Kt} = u_{it} \tag{4.8}$$

pour tout $i = 1 \ldots G$, et pour tout $t = 1 \ldots n$, où les Y_j sont les G variables endogènes et les X_l sont les K variables prédéterminées. Souvent, dans les équations, le coefficient de l'un des Y_j est normalisé à 1 (par exemple $\beta_{ii} = 1$ pour tout $i = 1 \ldots G$) et la variable explicative X_{1t} vaut 1, pour tout t, de sorte qu'elles peuvent s'écrire :

$$Y_{it} = -\gamma_{i1} + \sum_{l \neq i} (-\beta_{il})\, Y_{lt} + (-\gamma_{i2})\, X_{2t} + \cdots + (-\gamma_{iK})\, X_{Kt} + u_{it} \tag{4.9}$$

Dans la suite du chapitre, on utilisera la formulation précédente.

Des **identités** qui relient certaines variables du modèle, sans faire intervenir des coefficients inconnus et sans terme d'erreur, remplacent parfois certaines équations structurelles. Si on a G_1 équations structurelles et G_2 identités, $G = G_1 + G_2$. Il faut G équations ou identités pour pouvoir déterminer G variables endogènes.

Une équation structurelle ne fait pas nécessairement intervenir toutes les variables : parmi les endogènes et les prédéterminées, certaines peuvent être absentes (leurs coefficients peuvent être nuls). Des restrictions *a priori* peuvent en outre relier certains coefficients. Généralement, on pose la condition que la première variable prédéterminée x_1 est exogène et vaut 1 à toute période t, de manière à avoir une constante dans les équations. Le système formé par les équations structurelles et les identités s'appelle **forme structurelle** et peut s'exprimer de manière matricielle :

$$By_t + \Gamma x_t = u_t \tag{4.10}$$

pour tout $t = 1 \ldots n$, où $y_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{Gt} \end{pmatrix}$, $x_t = \begin{pmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{Kt} \end{pmatrix}$, $u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Gt} \end{pmatrix}$, $B = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1G} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{G1} & \beta_{G2} & \cdots & \beta_{GG} \end{pmatrix}$ et $\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{G1} & \gamma_{G2} & \cdots & \gamma_{GK} \end{pmatrix}$. La matrice B a G colonnes et G lignes. La matrice Γ a G lignes et K colonnes. Si l'équation i est une identité, $u_{it} = 0$ et les coefficients β_{ij} et γ_{ij} ont des valeurs connues dès le départ. Si toutes les équations ou identités sont normalisées, $\beta_{ii} = 1$ pour tout i. Si une constante figure dans certaines équations, $X_{1t} = 1$ pour tout t.

La théorie économique, financière, ou marketing, suggère des restrictions *a priori* sur les paramètres de B et Γ. Sans ces restrictions, les G équations seraient similaires et il serait impossible d'estimer leurs coefficients. Quand on fixe les valeurs des éléments de B et G, la forme structurelle est un système de G équations permettant de déterminer les valeurs des G variables endogènes Y_{it} en fonction des K variables prédéterminées X_{jt}. Si la matrice B n'est pas singulière, la solution du système s'appelle **forme réduite** et s'exprime de la façon suivante :

$$y_t = -B^{-1}\Gamma x_t + B^{-1}u_t = \Pi x_t + v_t \tag{4.11}$$

où $\Pi = -B^{-1}\Gamma$ et $v_t = B^{-1}u_t$.



Si l'on multiplie le système $By_t + \Gamma x_t = u_t$ par une matrice F non singulière à G lignes et G colonnes, on obtient un nouveau système équivalent : $FBy_t + F\Gamma x_t = Fu_t$. Si les hypothèses sur la distribution des u_{it} restent inchangées, on montre que le nouveau système et l'ancien ont exactement la même fonction de vraisemblance. Il manque donc des informations dans les données pour déterminer si les coefficients « vrais inconnus » valent B et Γ, ou bien FB et $F\Gamma$. Les deux formes structurelles sont équivalentes en observation. On montre qu'elles ont exactement la même forme réduite. Sans restrictions, les coefficients de B et Γ ne sont donc pas identifiés : plusieurs valeurs de B et Γ impliquent les mêmes valeurs des coefficients Π de la forme réduite. Si l'on connaît les valeurs des coefficients de Π, on ne peut donc pas trouver de manière unique les coefficients de B et Γ. Pour identifier certains paramètres de B et Γ, il faut imposer des restrictions valables à la fois sur les coefficients des matrices B et Γ et sur les coefficients des matrices FB et $F\Gamma$ résultant de toute transformation de la forme structurelle.

Les coefficients d'une équation structurelle sont identifiés quand seules des valeurs uniques, compatibles avec des valeurs fixées des coefficients de la forme réduite, sont en jeu. Par conséquent, les restrictions posées *a priori* sur B et Γ, et qui doivent être conservées par des transformations FB et $F\Gamma$, doivent impliquer suffisamment de restrictions sur toute matrice de transformations F pour que les coefficients contenus dans FB et $F\Gamma$ soient identiques à ceux contenus dans B et Γ.

Le lien entre les coefficients de la forme structurelle et les coefficients de la forme réduite est bien résumé par la relation $AW = 0$ où $A = [B\ \Gamma]$ et $W = \begin{bmatrix} \Pi \\ I_K \end{bmatrix}$. La matrice A a G lignes et $(G+K)$ colonnes. La matrice W a $(G+K)$ lignes et G colonnes. Chaque ligne de A est notée α_i, pour $i = 1 \ldots G$. Une telle ligne est une matrice à 1 ligne et $(G+K)$ colonnes. Évidemment, $\alpha_i W = 0$ pour tout $i = 1 \ldots G$. Imposer r_i restrictions aux coefficients de α_i implique qu'il existe une matrice φ_i, avec $(G+K)$ lignes et r_i colonnes, telle que $\alpha_i \varphi_i = 0$. La ligne α_i doit donc respecter la condition générale $\alpha_i[W\ \ \varphi_i] = 0$. La matrice $[W\ \ \varphi_i]$ a $(G+K)$ lignes et $(G+r_i)$ colonnes. Si, en plus, on normalise α_i en posant qu'un de ses coefficients est égal à 1, il faut, pour que l'équation $\alpha_i[W\ \ \varphi_i] = 0$ ait une solution α_i unique, que le rang de $[W\ \ \varphi_i]$ soit égal à $G+K-1$, ce qui est réalisé si et seulement si le rang de $A\varphi_i$ est égal à $G-1$. Il faut pour cela que r_i soit supérieur ou égal à $G-1$. En d'autres termes, on peut identifier les coefficients d'une équation structurelle si le nombre de restrictions *a priori* sur lesdits coefficients est au moins égal à $G-1$, où G est le nombre d'équations et identités du système (ou de variables endogènes). Quand seules des restrictions d'exclusion portant sur les variables de l'équation sont posées, il faut, pour opérer cette identification, que le nombre r_i de variables exclues soit au moins égal à $G-1$, c'est-à-dire que le nombre de variables prédéterminées exclues soit au moins égal à celui des variables endogènes présentes dans l'équation moins une.

Une équation est « exactement identifiée » quand le nombre de restrictions est égal au minimum requis pour l'identification ; quand il est supérieur à ce minimum, on dit qu'elle est « suridentifiée ».

3 Méthode des doubles moindres carrés

Abstraction faite d'éventuelles identités, chaque équation structurelle se présente encore de la manière suivante :

$$Y_i = Y^i\beta_i + X_i\gamma_i + u_i \tag{4.12}$$

pour tout $i = 1 \dots G_1$. En l'absence d'identités, $G_1 = G$. La notation Y_i représente le vecteur des n observations de la i^e variable endogène. Les vecteurs d'observation des g_i autres variables endogènes présentes dans l'équation i forment les colonnes de la matrice Y^i, à n lignes et g_i colonnes. Les vecteurs d'observation des k_i variables prédéterminées forment la matrice X_i, à n lignes et k_i colonnes. Bien entendu, $g_i \leqslant G - 1$ et $k_i \leqslant K$. Le vecteur colonne β_i est composé des valeurs non nulles $-\beta_{ij}$ ($j = 1 \dots G$ et $j \neq i$), où les β_{ij} sont les éléments non nuls de la ligne i de la matrice B (dans laquelle les β_{ii} valent 1). Le vecteur colonne γ_i est formé des valeurs non nulles $-\gamma_{ij}$ ($j = 1 \dots K$), où les γ_{ij} sont les éléments non nuls de la ligne i de la matrice Γ. Chaque équation i se représente encore de la manière suivante :

$$Y_i = (Y^i X_i) \begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix} + u_i = Z_i \lambda_i + u_i \tag{4.13}$$

où $Z_i = (Y^i X_i)$ est une matrice à n lignes et $(g_i + k_i)$ colonnes, et où $\begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix} = \lambda_i$.

Le choix d'une méthode d'estimation d'un modèle structurel dépend évidemment des hypothèses faites sur la distribution du terme d'erreur. Généralement, on suppose que les termes d'erreur contemporains des différentes équations sont corrélés entre eux : $\text{Cov}(u_{it}, u_{jt}) = \sigma_{ij}$ pour tout $i = 1 \dots m$ et $j = 1 \dots m$, et que les termes d'erreur non contemporains (de périodes différentes) ne sont pas corrélés entre eux : $\text{Cov}(u_{it}, u_{jt-\theta}) = 0$ pour tout $i = 1 \dots m$ et $j = 1 \dots m$ et pour tout $\theta \geqslant 1$. Le terme d'erreur de chaque équation est supposé homoscédastique et non autocorrélé : $\text{Cov}(u_{it}, u_{it-\theta}) = 0$ et $V(u_{it}) = \sigma_i^2$ pour tout $i = 1 \dots m$. Ces hypothèses impliquent que :

$$\Sigma_{u_i u_j} = E(u_i u_j') = \begin{pmatrix} \sigma_{ij} & 0 & \cdots & 0 \\ 0 & \sigma_{ij} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{ij} \end{pmatrix} = \sigma_{ij} I_n$$

et

$$\Sigma_{u_i} = E(u_i u_i') = \begin{pmatrix} \sigma_i^2 & 0 & \cdots & 0 \\ 0 & \sigma_i^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_i^2 \end{pmatrix} = \sigma_i^2 I_n \tag{4.14}$$

On peut néanmoins choisir d'ignorer les liens entre les termes d'erreur des différentes équations, et estimer chacune d'entre elles individuellement. Ce faisant, on produit des estimateurs moins précis que ceux obtenus lorsque l'on tient compte de la corrélation contemporaine entre les termes d'erreur des différentes équations. Toutefois, de tels estimateurs peuvent être au moins non biaisés ou convergents, à condition que l'on tienne compte de la simultanéité. Effectivement, la présence de variables endogènes comme variables explicatives d'autres variables endogènes implique que le terme d'erreur de chaque équation n'est généralement pas indépendant de toutes les variables explicatives de cette équation (voir chapitre 2). Par conséquent, l'estimation de chaque équation par la méthode des MCO est exclue : on obtiendrait des estimateurs biaisés et non convergents. Il faut donc recourir à une méthode de variables instrumentales.



La méthode des doubles moindres carrés, dite 2MC, notée 2SLS en anglais (2 Stages Least Squares), est l'estimation séparée de chaque équation du système par la méthode des variables instrumentales, qui sont en l'occurrence toutes les variables prédéterminées du système (au-delà de l'équation i) qui forment la matrice X_T telle que :

$$X_T = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{K1} \\ X_{12} & X_{22} & \cdots & X_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{Kn} \end{pmatrix}$$

L'estimateur des doubles moindres carrés de λ_i est donc :

$$\hat{\lambda}_i^{2MC} = \left(Z_i' X_T (X_T' X_T)^{-1} X_T' Z_i\right) Z_i X_T (X_T' X_T)^{-1} X_T' Y_i \tag{4.15}$$

On peut aussi obtenir cet estimateur en deux étapes :

1) On estime une régression de chaque variable Y_j^i de Y^i $(j = 1 \ldots g_i)$ sur toutes les variables prédéterminées du système, par la méthode des moindres carrés ordinaires : $Y_j^i = X_T \hat{w}_j^{i\,MCO} + e_j$ avec $\hat{w}_j^{i\,MCO} = \left(X_T' X_T\right)^{-1} (X_T)' \, Y_j^i$, et l'on garde les valeurs calculées ainsi obtenues : $\hat{Y}_j^i = X \hat{w}_j^{i\,MCO} = X_T \left(X_T' X_T\right)^{-1} (X_T)' \, Y_j^i$. On les rassemble dans une matrice $\hat{Y}^i = \left(\hat{Y}_1^i \hat{Y}_2^i \cdots \hat{Y}_{gi}^i\right) = X_T \left(X_T' X_T\right)^{-1} (X_T)' \, Y^i$. On définit alors la matrice $\hat{Z}_i = \left(\hat{Y}^i \; X_i\right)$.
2) On estime une régression de Y_i sur les variables $\hat{Y}_j^i$ et sur les variables de X_i par MCO. Il s'agit donc d'une régression de Y_i sur les variables de $\hat{Z}_i = \left(\hat{Y}^i \; X_i\right)$: $Y_i = \hat{Z}_i \hat{\lambda}_i^{2MC} + E_i$. où

$$\hat{\lambda}_i^{2MC} = \left(\hat{Z}_i' \hat{Z}_i\right)^{-1} \hat{Z}_i' Y_i \tag{4.16}$$

On peut encore présenter l'estimateur des doubles moindres carrés en agrégeant les différentes équations de la manière suivante :

$$Y = Z\lambda + u \tag{4.17}$$

où $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{G_1} \end{pmatrix}$, $Z = \begin{pmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_{G_1} \end{pmatrix}$, $\hat{\lambda}^{2MC} = \begin{pmatrix} \hat{\lambda}_1^{2MC} \\ \lambda_2^{2MC} \\ \vdots \\ \lambda_{G_1}^{2MC} \end{pmatrix}$ et $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{G_1} \end{pmatrix}$.

La matrice de variance et de covariance de u est :

$$\Sigma_u = E(uu_i') = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_m^2 \end{pmatrix} \otimes I_n = \Sigma \otimes I_n$$

Après projection des variables endogènes de droite sur les variables instrumentales (toutes les variables endogènes du modèle), on peut définir le modèle agrégé estimé suivant :

$$Y = \hat{Z}\hat{\lambda}^{2MC} + E \qquad (4.18)$$

où $\hat{Z} = \begin{pmatrix} \hat{Z}_1 & 0 & \cdots & 0 \\ 0 & \hat{Z}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{Z}_{G_1} \end{pmatrix}$, $\hat{\lambda}^{2MC} = \begin{pmatrix} \hat{\lambda}_1^{2MC} \\ \hat{\lambda}_2^{2MC} \\ \vdots \\ \hat{\lambda}_{G_1}^{2MC} \end{pmatrix}$ et $E = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_{G_1} \end{pmatrix}$, et où

$$\hat{\lambda}^{2MC} = \left(\hat{Z}'\hat{Z}\right)^{-1}\hat{Z}'Y = \left(\hat{Z}'Z\right)^{-1}\hat{Z}'Y \qquad (4.19)$$

$\hat{B}^{2MC}$ et $\hat{\Gamma}^{2MC}$ sont donc les estimateurs de B et Γ. Une estimation de la matrice de variance et de covariance de $\hat{\lambda}_i^{2MC}$ est donnée par :

$$\begin{aligned}\hat{\Sigma}_{\hat{\lambda}_i^{2MC}} &= \left(\hat{Z}_i'\hat{Z}_i\right)^{-1}\frac{\left(Y_i - Z_i\hat{\lambda}_i^{2MC}\right)\left(Y_i - Z_i\hat{\lambda}_i^{2MC}\right)'}{n} \\ &= \left(Z_i'\hat{Z}_i\right)^{-1}\frac{\left(Y_i - Z_i\hat{\lambda}_i^{2MC}\right)\left(Y_i - Z_i\hat{\lambda}_i^{2MC}\right)'}{n}\end{aligned}$$

On montre que les estimateurs des doubles moindres carrés sont convergents. Toutefois, ils présentent un biais de petit échantillon ; ils sont asymptotiquement normaux. Comme ils ne tiennent pas compte des corrélations entre les termes d'erreur (ou des perturbations) des différentes équations, ils sont en général asymptotiquement inefficients, au sens où leur précision est moindre que celle des estimateurs que l'on obtient en tenant compte de ces corrélations. En contrepartie de cette insuffisance, ils présentent l'avantage ne pas être affectés par d'éventuelles erreurs de spécification dans les différentes autres équations puisqu'ils n'en tiennent pas compte.

4 Méthode des triples moindres carrés

La méthode des triples moindres carrés, dite 3MC, notée en anglais 3SLS (3 Stages Least Squares), est à information complète. L'estimateur des triples moindres carrés est l'estimateur 2MC avec une correction du type des moindres carrés généralisés, qui tient compte des corrélations contemporaines entre les termes d'erreur des équations structurelles du modèle.

Il s'agit donc d'une nouvelle manière d'estimer les G_1 équations (4.12), qui peuvent être agrégées en la formule (4.17), et dont les termes d'erreur contemporains sont corrélés, comme indiqué par la matrice Σ. On réalise les étapes 1 et 2 de la méthode des doubles moindres carrés, puis on utilise les résidus E_i ainsi obtenus pour estimer les éléments de la matrice Σ :

$$\hat{\sigma}_i^2 = \frac{E_i'E_i}{n - k_i} \quad \text{et} \quad \hat{\sigma}_{ij} = \frac{E_i'E_j}{\sqrt{n - k_i}\sqrt{n - k_i j}} \qquad (4.20)$$

Ces estimations forment une matrice $\hat{\Sigma}$ estimée, qui peut être utilisée dans la formule des MCG. On obtient ainsi les estimateurs 3MC des coefficients :

$$\hat{\lambda}^{3MC} = \left(\hat{Z}'\left(\hat{\Sigma}^{-1} \otimes I_n\right)\hat{Z}\right)^{-1}\hat{Z}'\left(\hat{\Sigma}^{-1} \otimes I_n\right)Y \qquad (4.21)$$



$\hat{B}^{3MC}$ et $\hat{\Gamma}^{3MC}$ sont donc les estimateurs de B et Γ. Lorsque toutes les équations sont exactement identifiées, les estimateurs des triples moindres carrés sont identiques aux estimateurs des doubles moindres carrés : $\hat{\lambda}^{3MC} = \hat{\lambda}^{2MC}$.

Une estimation de la matrice de variance et de covariance de $\hat{\lambda}^{3MC}$ est donnée par :

$$\hat{\Sigma}_{\hat{\lambda}^{3MC}} = \left(\hat{Z}'\left(\hat{\Sigma}^{-1} \otimes I_n\right)\hat{Z}\right)^{-1} \qquad (4.22)$$

On constate que les estimateurs des triples moindres carrés sont convergents. Toutefois, ils présentent un biais de petit échantillon ; ils sont asymptotiquement normaux. Lorsque certaines équations du modèle sont suridentifiées, ils sont plus efficients que les estimateurs 2MC.

5 Méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance, notée FIML (Full Information Maximum Likelihood), est à information complète. Contrairement aux méthodes 2MC et 3MC, elle nécessite de faire des hypothèses explicites sur la distribution. En général, on suppose que le modèle structurel $By_t + \Gamma x_t = u_t$ est tel que $u_t \sim N(0, \Sigma)$.

Pour $G_1 = G$, cette distribution normale du terme d'erreur implique que la fonction de densité de y_t, conditionnellement à x_t, est donnée par :

$$(2\pi)^{-\frac{G}{2}} |\det B| \left(\det \sum\right)^{-\frac{1}{2}} e^{-\frac{1}{2}(By_t+\Gamma x_t)' \sum^{-1}(By_t+\Gamma x_t)} \qquad (4.23)$$

L'hypothèse d'indépendance intertemporelle implique alors que le logarithme de la fonction de vraisemblance est :

$$\begin{aligned}\ln\bigl(L(B, \Gamma, \Sigma)\bigr) = &-\frac{Gn}{2}\ln(2\pi) + n\ln(|\det B|)\\ &-\frac{n}{2}\ln\left(\det\sum\right) - \frac{1}{2}\sum_{t=1}^{n}\left(By_t + \Gamma x_t\right)' \sum^{-1}\left(By_t + \Gamma x_t\right)\end{aligned} \qquad (4.24)$$

Les estimateurs de maximum de vraisemblance B^{MV}, Γ^{MV} et Σ^{MV} sont les valeurs de B, Γ et Σ pour lesquelles cette fonction est maximale. Ils sont convergents et asymptotiquement efficients, à condition que le choix de la distribution soit correct. Ils sont asymptotiquement distribués selon une normale.

Résumé

Les systèmes de régressions simultanées regroupent des équations qui expliquent chacune une variable endogène, en fonction de variables explicatives différentes, qui ne comportent aucune variable endogène. Les perturbations ou termes d'erreur sont indépendants des variables explicatives. Les équations sont liées pour la simple raison que les termes d'erreur contemporains sont corrélés entre eux. La méthode SUR permet d'obtenir des estimateurs plus précis que ceux que l'on a en appliquant les *MCO* à chaque équation. Les modèles à équations simultanées regroupent des équations qui expliquent chacune une variable endogène en fonction de variables prédéterminées et souvent aussi d'autres variables endogènes, qui sont elles-mêmes expliquées par d'autres équations du modèle. Cette simultanéité implique que les termes d'erreur sont liés aux variables explicatives. On ne peut estimer les coefficients que sous certaines conditions d'identification. La méthode des doubles moindres carrés, qui est à information limitée, et celles des triples moindres carrés et du maximum de vraisemblance, qui sont à information complète, sont des outils d'estimation appropriés.



Problèmes et exercices

EXERCICE 1 MODÈLE MACROÉCONOMIQUE ET MÉTHODE 2MC

Énoncé

Le fichier FR.xls, téléchargeable sur le site Internet afférent à cet ouvrage, contient les variables suivantes, sur une période allant de 1978 à 2002 :

- CF : consommation des ménages à prix constants ;
- I : investissement ou formation brute de capital fixe à prix constants ;
- X : exportations à prix constants ;
- M : importations à prix constants ;
- PIB : produit intérieur brut à prix constants ;
- RPA : revenu primaire réel des autres secteurs que les ménages ;
- $PRELN$: prélèvements d'impôts et cotisations sociales sur les ménages en termes réels, nets de prestations reçues par les ménages ;
- P : indice des prix de la consommation ;
- PM : indice des prix des importations.

Sur la base de ces données, on peut définir les variables suivantes :

- RDR : revenu disponible réel des ménages, égal à $RPR - PRELN$;
- RES : autres composantes de la demande domestique à prix constants, égales à $PIB - (CF + I + X - M)$;
- RPR : revenu primaire réel des ménages, égal à $PIB - RPA$;
- PMR : prix relatif des importations, égal à PM/P ;
- INF : taux d'inflation égal à la variation logarithmique de P.

On spécifie un modèle macroéconomique dont les équations structurelles sont les suivantes :

$$\begin{aligned}
CF_t &= \alpha_1 + \alpha_2 RDR_t + \alpha_3 RDR_{t-1} + \alpha_4 CF_{t-1} + \alpha_5 INF_t + u_{1t}\\
I_t &= \beta_1 + \beta_2 PIB_t + \beta_3 PIB_{t-1} + \beta_4 I_{t-1} + u_{2t}\\
M_t &= \gamma_1 + \gamma_2 PIB_t + \gamma_3 PMR_t + \gamma_4 M_{t-1} + u_{3t}
\end{aligned}$$

Les identités sont :

$$\begin{aligned}
PIB_t &= CF_t + I_t + RES_t + X_t - M_t\\
RPR_t &= PIB_t - RPA_t\\
RDR_t &= RPR_t - PRELN_t
\end{aligned}$$

Les variables endogènes sont CF_t, RDR_t, $I_t PIB_t$, M_t et RPR_t. Les variables exogènes sont INF_t et PMR_t. Les variables prédéterminées sont CF_{t-1}, RDR_{t-1}, I_{t-1}, PIB_{t-1}, INF_t, PMR_t et M_{t-1}. Estimez ce modèle par la méthode des doubles moindres carrés. Travaillez avec TSP.

Solution

Il faut d'abord vérifier que les conditions d'identification de chaque équation sont respectées. La première équation contient deux variables endogènes : CF_t et RDR_t, et en exclut quatre prédéterminées : I_{t-1}, PIB_{t-1}, PMR_t et M_{t-1}. L'équation est identifiée, et même suridentifiée, puisque $4 > 2-1$. La deuxième équation contient deux variables endogènes : I_t et PIB_t, et en exclut cinq prédéterminées : CF_{t-1}, RDR_{t-1}, INF_t, PMR_t et M_{t-1}. L'équation est identifiée, et même suridentifiée, puisque $5 > 2-1$. La troisième équation contient deux variables endogènes : M_t et PIB_t, et en exclut cinq prédéterminées : CF_{t-1}, RDR_{t-1}, I_{t-1}, PIB_{t-1} et INF_t L'équation est identifiée, et même suridentifiée, puisque $5 > 2-1$.

Le programme d'instructions TSP est le suivant :

```
OPTIONS MEMORY=25;
FREQ A;
SMPL 1978 2002;
READ(FILE='C :\FRC.XLS');
PMR=PM/P;
RES=PIB-(CF+I+X-M);
RPR=PIB-RPA;
RDR=RPR-PRELN;
SMPL 1979 2002;
INF=LOG(P)-LOG(P(-1));
2SLS(INST=(C,CF(-1),INF,PIB(-1) I(-1) PMR M(-1) RPÀ PRELN)) CF C RDR RDR(-1) CF(-1) INF;
2SLS(INST=(C,CF(-1),INF,PIB(-1) I(-1) PMR M(-1) RPÀ PRELN)) I C PIB PIB(-1) I(-1);
2SLS(INST=(C,CF(-1),INF,PIB(-1) I(-1) PMR M(-1) RPÀ PRELN)) M C PIB PMR M(-1);
```

Les résultats sont :

```
                              Equation 1
                             ============

                Method of estimation = Instrumental Variable

Dependent variable: CF
Endogenous variables: RDR RDR(-1)
Included exogenous variables: C CF(-1) INF
Excluded exogenous variables: PIB(-1) I(-1) PMR M(-1) RP À PRELN
Current sample:  1979 to 2002
Number of observations:  24

       Mean of dep. var. = 613.704           R-squared = .996999
  Std. dev. of dep. var. = 75.2243  Adjusted R-squared = .996368
Sum of squared residuals = 390.795       Durbin-Watson = 2.89278 [.541,1.00]
   Variance of residuals = 20.5682              E'PZ*E = 125.497
Std. error of regression = 4.53521

           Estimated    Standard
Variable  Coefficient     Error       t-statistic    P-value
C         75.3664       19.9505       3.77767        [.000]
RDR       .889724       .148381       5.99621        [.000]
RDR(-1)   -.616957      .147234       -4.19030       [.000]
```



```
CF(-1)    .566376      .144285      3.92541       [.000]
INF       -182.553     63.2190      -2.88762      [.004]

                          Equation 2
                         ============

              Method of estimation = Instrumental Variable

Dependent variable: I
Endogenous variables: PIB
Included exogenous variables: C PIB(-1) I(-1)
Excluded exogenous variables: CF(-1) INF PMR M(-1) RPÀ PRELN
Current sample:  1979 to 2002
Number of observations:  24

       Mean of dep. var. = 214.343          R-squared = .992031
  Std. dev. of dep. var. = 36.6089  Adjusted R-squared = .990835
Sum of squared residuals = 246.240      Durbin-Watson = 1.93225 [.002,.992]
   Variance of residuals = 12.3120             E'PZ*E = 40.9836
Std. error of regression = 3.50884

           Estimated     Standard
Variable  Coefficient     Error       t-statistic    P-value
C         -3.80473       5.21909       -.729004      [.466]
PIB       .642299        .059717       10.7558       [.000]
PIB(-1)   -.623924       .065328       -9.55064      [.000]
I(-1)     .875597        .075301       11.6280       [.000]

                          Equation 3
                         ============

             Method of estimation = Instrumental Variable

Dependent variable: M
Endogenous variables: PIB
Included exogenous variables: C PMR M(-1)
Excluded exogenous variables: CF(-1) INF PIB(-1) I(-1) RPA
    PRELN
Current sample:  1979 to 2002
Number of observations:  24

       Mean of dep. var. = 221.583          R-squared = .984135
  Std. dev. of dep. var. = 78.6838  Adjusted R-squared = .981755
Sum of squared residuals = 2259.10      Durbin-Watson = 1.52590 [.000,.880]
   Variance of residuals = 112.955             E'PZ*E = 1614.40
Std. error of regression = 10.6280

           Estimated      Standard
Variable  Coefficient      Error       t-statistic   P-value
C         -79.5457       93.2394       -.853133      [.394]
PIB       .117524        .083857       1.40149       [.161]
PMR       1.41185        25.9975       .054307       [.957]
M(-1)     .804479        .143033       5.62442       [.000]
```

EXERCICE 2 MODÈLE MACROÉCONOMIQUE ET MÉTHODE 3MC

Énoncé

Sur la base des données de l'exercice 1, estimez le même modèle que celui de l'exercice précédent, en utilisant cette fois la méthode des triples moindres carrés. Travaillez avec TSP.

Solution

Le programme d'instructions TSP est le suivant :

```
OPTIONS MEMORY=25;
FREQ A;
SMPL 1978 2002;
READ(FILE='C :\FR.XLS');
PMR=PM/P;
RES=PIB-(CF+I+X-M);
RPR=PIB-RPA;
RDR=RPR-PRELN;
SMPL 1979 2002;
INF=LOG(P)-LOG(P(-1));
FORM(VARPREF=A) CON CF C RDR RDR(-1) CF(-1) INF;
FORM(VARPREF=B) INV I C PIB PIB(-1) I(-1);
FORM(VARPREF=C) IMP M C PIB PMR M(-1);
3SLS(INST=(C,CF(-1),INF,PIB(-1) I(-1) PMR M(-1) RPÀ PRELN)) CON INV IMP;
```

Les résultats sont :

```
                    THREE STAGE LEAST SQUARES
                    =========================

EQUATIONS: CON INV IMP

INSTRUMENTS: C CF(-1) INF PIB(-1) I(-1) PMR M(-1) RPÀ PRELN

MAXIMUM NUMBER OF ITERATIONS ON V--COV MATRIX OF RESIDUALS = 0
NOTE => The model is linear in the parameters.
Working space used: 5025

                        STARTING VALUES

                A0         ARDR        ARDR1        ACF1        AINF
VALUE      0.00000      0.00000      0.00000     0.00000     0.00000

                B0         BPIB        BPIB1         BI1          C0
VALUE      0.00000      0.00000      0.00000     0.00000     0.00000

              CPIB         CPMR          CM1
VALUE      0.00000      0.00000      0.00000


F= 0.11622E+08 FNEW= 1780.9 ISQZ= 0 STEP= 1.0000 CRIT= 0.11621E+08

CONVERGENCE ACHIEVED AFTER 1 ITERATION
```




```
 2 FUNCTION EVALUATIONS.


END OF TWO STAGE LEAST SQUARES ITERATIONS (SIGMA=IDENTITY). THREE STAGE
LEAST SQUARES ESTIMATES WILL BE OBTAINED USING THIS ESTIMATE OF SIGMA:

                       RESIDUAL COVARIANCE MATRIX

                    CON           INV           IMP
CON           16.28313
INV           0.072019      10.25999
IMP            2.37859      -1.17726      94.12937

                            WEIGHTING MATRIX

                    CON            INV            IMP

CON            0.24782     -0.0013808      -0.015148
INV                           0.31220       0.011964
IMP                                         0.10334

Working space used: 5025
F=   28.337    FNEW=   28.284   ISQZ=   0 STEP=  1.0000   CRIT= 0.53014E-01

CONVERGENCE ACHIEVED AFTER 1 ITERATION


 4 FUNCTION EVALUATIONS.


                          THREE STAGE LEAST SQUARES
                          =========================

                          Residual Covariance Matrix

                    CON            INV            IMP
CON           16.07974
INV          -0.017320       10.37265
IMP            3.45061       -2.00440       94.29357

                              Weighting Matrix

                    CON            INV            IMP

CON            0.24782      -0.0013808     -0.015148
INV                            0.31220      0.011964
IMP                                          0.10334

                 Covariance Matrix of Transformed Residuals

                    CON             INV            IMP
CON           23.70022
INV           -0.16422        24.26539
IMP            0.67084        -0.62395      23.91249
```

Exercices

```
Number of observations = 24 E'PZ*E = 28.2839
```

Parameter	Estimate	Standard Error	t-statistic	P-value
A0	73.4501	17.7272	4.14337	[.000]
ARDR	.872145	.131794	6.61746	[.000]
ARDR1	-.610985	.130769	-4.67223	[.000]
ACF1	.583227	.128190	4.54971	[.000]
AINF	-177.335	56.1668	-3.15729	[.002]
B0	-3.84120	4.76406	-.806288	[.420]
BPIB	.648311	.054476	11.9009	[.000]
BPIB1	-.629440	.059594	-10.5621	[.000]
BI1	.872549	.068693	12.7022	[.000]
C0	-72.2268	84.9437	-.850291	[.395]
CPIB	.113072	.076427	1.47948	[.139]
CPMR	-1.05115	23.6783	-.044393	[.965]
CM1	.807163	.130411	6.18939	[.000]

```
Standard Errors computed from quadratic form of analytic first
derivatives (Gauss)

Equation: CON
Dependent variable: CF

       Mean of dep. var. = 613.704
  Std. dev. of dep. var. = 75.2243
Sum of squared residuals = 385.914
   Variance of residuals = 16.0797
Std. error of regression = 4.00996
               R-squared = .997037
           Durbin-Watson = 2.87331 [.517,1.00]

Equation: INV
Dependent variable: I

       Mean of dep. var. = 214.343
  Std. dev. of dep. var. = 36.6089
Sum of squared residuals = 248.944
   Variance of residuals = 10.3727
Std. error of regression = 3.22066
               R-squared = .991944
           Durbin-Watson = 1.95254 [.003,.994]

Equation: IMP
Dependent variable: M

       Mean of dep. var. = 221.583
 Std.  dev. of dep. var. = 78.6838
Sum of squared residuals = 2263.05
   Variance of residuals = 94.2936
Std. error of regression = 9.71049
               R-squared = .984108
           Durbin-Watson = 1.53330 [.000,.884]
```



Exercice 3 Modèle macroéconomique et méthode MV

Énoncé

On spécifie un modèle macroéconomique dont les équations structurelles sont les suivantes :

$$\ln(CF_t) - \ln(CF_{t-1}) = \alpha_1 + \alpha_2\big(\ln(RDR_t) - \ln(RDR_{t-1})\big) + \alpha_3 \ln(RDR_{t-1})$$
$$+ \alpha \ln(CF_{t-1}) + \alpha(\ln(P_t) - \ln(P_{t-1})) + u_{1t}$$
$$\ln(I_t) - \ln(I_{t-1}) = \alpha_1 + \alpha_2\big(\ln(PIB_t) - \ln(PIB_{t-1})\big) + \alpha_3 \ln(I_{t-1}) + \alpha \ln(PIB_{t-1})$$
$$+ \alpha \ln(CF_{t-1}) + \alpha\big(\ln(P_t) - \ln(P_{t-1})\big) + u_{2t}$$
$$\ln(M_t) - \ln(M_{t-1}) = \alpha_1 + \alpha_2 \ln PIB_t - \ln PIB_{t-1} + \alpha_3 \ln(PMR_t) - \ln(PMR_{t-1})$$
$$+ \alpha \ln(PIB_{t-1}) + \alpha \ln(PMR_{t-1}) + \alpha \ln(M_{t-1}) + u_{3t}$$

Les identités sont :

$$PIB_t = CF_t + I_t + RES_t + X_t - M_t$$
$$RPR_t = PIB_t - RPA_t$$
$$RDR_t = RPR_t - PRELN_t$$

Sur la base des données de l'exercice 1, estimez ce modèle macroéconomique par la méthode FIML. Travaillez avec TSP.

Solution

Le programme d'instructions TSP est le suivant :

```
OPTIONS MEMORY=25;
FREQ A;
SMPL 1978 2002;
READ(FILE='C :\FRC.XLS');
LP=LOG(P);
LPM=LOG(PM);
PMR=PM/P;
LPMR=LPM-LP;
LPB=LOG(PIB);
LM=LOG(M);
LPB=LOG(PIB);
RES=PIB-(CF+I+X-M);
RPR=PIB-RPA ;
RDR=RPR-PRELN;
LI=LOG(I);
LC=LOG(CF);
LY=LOG(RDR);
SMPL 1979 2002;
DLC=LC-LC(-1);
DLY=LY-LY(-1);
DLPMR=LOG(PMR)-LOG(PMR(-1));
DLM=LM-LM(-1) ;
DLP=LOG(P)-LOG(P(-1));
```

```
DLPIB=LPB-LPB(-1);
DLI=LI-LI(-1);
PARAM A1 A2 A3 A4 A5 B1 B2 B3 B5 B4 C1 C2 C4 C5 C3 C6;
OLSQ DLC C DLY LY(-1) LC(-1) DLP;
SET A1=@CoeF(1);
SET A2=@CoeF(2);
SET A3=@CoeF(3);
SET A4=@CoeF(4);
SET A5=@CoeF(5);
OLSQ DLI C DLPIB LI(-1) LPB(-1) DLP;
SET B1=@CoeF(1);
SET B2=@CoeF(2);
SET B3=@CoeF(3);
SET B4=@CoeF(4);
OLSQ DLM C DLPIB DLPMR LPMR(-1) LPB(-1) LM(-1);
SET C1=@CoeF(1);
SET C2=@CoeF(2);
SET C3=@CoeF(3);
SET C4=@CoeF(4);
SET C5=@CoeF(5);
SET C6=@CoeF(6);
IDENT IDRDR RDR=RPR-PRELN;
IDENT IDDLC DLC=LC-LC(-1);
IDENT IDDLY DLY=LY-LY(-1);
IDENT ILC LC=LOG(CF);
IDENT ILY LY=LOG(RDR);
IDENT IDDLI DLI=LI-LI(-1);
IDENT IDLPB LPB=LOG(PIB);
IDENT IDLPIB DLPIB=LPB-LPB(-1);
IDENT IDPIB PIB=CF+I-M+X+RES;
IDENT IDRPR RPR=PIB-RPA;
IDENT IDDLM DLM=LM-LM(-1);
IDENT ILM LM=LOG(M);
IDENT ILI LI=LOG(I);
FRML EQC DLC=A1+A2*DLY+A3*LY(-1)+A4*LC(-1)+A5*DLP;
FRML EQI DLI=B1+B2*DLPIB+B3*LI(-1)+B4*LPB(-1)+B5*DLP;
FRML EQM DLM=C1+C2*DLPIB+C3*DLPMR+C4*LPMR(-1)+C5*LPB(-1)+C6*LM(-1);
FIML(maxit=1000, HCOV=G, ENDOG=(CF,I,M,LC,LI,LM,DLC,DLI,DLM,RDR,
LY,DLY,RPR,LPB,DLPIB,PIB)) EQC EQI EQM IDRDR IDDLC IDDLY ILC ILY
IDLPB IDLPIB IDPIB IDRPR IDDLM ILM ILI IDDLI;
```

Il faut d'abord générer toutes les variables contenues dans les équations structurelles et les identités, à partir des données du fichier FR.xls. Ensuite, on définit les paramètres à estimer avec l'instruction `PARAM` : les `Ai` correspondent aux α_i, les `Bi` aux β_i et les `Ci` aux γ_i. On estime chaque équation du modèle par *MCO* et l'on utilise les valeurs estimées des coefficients, qui sont des valeurs de départ convenables, pour les paramètres à estimer par FIML. Après chaque estimation par *MCO*, TSP garde en mémoire le vecteur des coefficients estimés dans une variable appelée `@CoeF`. Les instructions `SET` affectent les valeurs successives de ce vecteur aux paramètres correspondants. Les instructions `IDENT`



définissent chaque identité en lui donnant un nom. Les instructions FRML définissent chaque équation structurelle en lui donnant un nom. L'instruction FIML demande l'estimation d'un système d'équations par la méthode du maximum de vraisemblance à information complète. Elle doit être suivie des noms de toutes les équations et identités : EQC, EQI, EQM, IDRDR, IDDLC, IDDLY, ILC, ILY, IDLPB, IDLPIB, IDPIB, IDRPR, IDDLM, ILM, ILI, IDDLI. L'option maxit=1000 limite TSP à un maximum de 1 000 itérations, dans ses tentatives d'obtenir la convergence des valeurs estimées. L'option HCOV=G requiert une méthode spécifique pour le calcul de la matrice de variance et de covariance estimées des coefficients estimés, qui détermine les valeurs des écarts types (ou standard errors) de ces derniers. La liste des variables endogènes doit être mentionnée : CF, I, M, LC, LI, LM, DLC, DLI, DLM, RDR, LY, DLY, RPR, LPB, DLPIB, PIB.

Les résultats de l'instruction FIML sont les suivants :

```
                    Information Maximum Likelihood
                  ===================================

Equations: EQC EQI EQM

Identities: IDDLI ILI ILM IDDLM IDRPR IDPIB IDLPIB IDLPB ILY
     ILC IDDLY IDDLC IDRDR

Endogenous variables: CF I M LC LI LM DLC DLI DLM RDR LY DLY
    RPR LPB DLPIB PIB

NOTE => The model is linear in the parameters.
Working space used: 9939

                           STARTING VALUES

                  A1            A2              A3              A4
VALUE        0.67685       0.81852         0.32262        -0.43377

                  A5            B1              B2              B3
VALUE       -0.34865     -0.061901         2.83055        -0.12893

                  B4            B5              C1              C2
VALUE        0.10334       0.00000        -1.03046         2.85777

                  C3            C4              C5              C6
VALUE      -0.048439      -0.19344         0.32068        -0.22365

F=   232.06    FNEW=   206.29    ISQZ=   3 STEP=  4.0000     CRIT=  20.690
F=   206.29    FNEW=   202.59    ISQZ=   1 STEP= 0.50000     CRIT=  15.929
F=   202.59    FNEW=   198.19    ISQZ=   2 STEP=  2.0000     CRIT=  6.4935
F=   198.19    FNEW=   198.10    ISQZ=   1 STEP=  1.0000     CRIT=  3.4928
F=   198.10    FNEW=   196.95    ISQZ=   1 STEP=  1.0000     CRIT=  2.4203
F=   196.95    FNEW=   196.88    ISQZ=   2 STEP=  2.0000     CRIT= 0.69374E-01
F=   196.88    FNEW=   196.86    ISQZ=   1 STEP=  1.0000     CRIT= 0.38891E-01
F=   196.86    FNEW=   196.84    ISQZ=   2 STEP=  2.0000     CRIT= 0.13301E-01
F=   196.84    FNEW=   196.83    ISQZ=   1 STEP=  1.0000     CRIT= 0.12607E-01
F=   196.83    FNEW=   196.83    ISQZ=   2 STEP=  2.0000     CRIT= 0.44010E-02
F=   196.83    FNEW=   196.83    ISQZ=   1 STEP=  1.0000     CRIT= 0.45739E-02
F=   196.83    FNEW=   196.83    ISQZ=   2 STEP=  2.0000     CRIT= 0.14973E-02
F=   196.83    FNEW=   196.82    ISQZ=   1 STEP=  1.0000     CRIT= 0.16746E-02
F=   196.82    FNEW=   196.82    ISQZ=   2 STEP=  2.0000     CRIT= 0.51582E-03
F=   196.82    FNEW=   196.82    ISQZ=   1 STEP=  1.0000     CRIT= 0.61612E-03
```

```
F=   196.82    FNEW=   196.82    ISQZ=   2 STEP=  2.0000     CRIT= 0.17904E-03
F=   196.82    FNEW=   196.82    ISQZ=   1 STEP=  1.0000     CRIT= 0.22777E-03
F=   196.82    FNEW=   196.82    ISQZ=   2 STEP=  2.0000     CRIT= 0.62537E-04
F=   196.82    FNEW=   196.82    ISQZ=   1 STEP=  1.0000     CRIT= 0.84587E-04
F=   196.82    FNEW=   196.82    ISQZ=   2 STEP=  2.0000     CRIT= 0.21967E-04
F=   196.82    FNEW=   196.82    ISQZ=   1 STEP=  1.0000     CRIT= 0.31545E-04
F=   196.82    FNEW=   196.82    ISQZ=   2 STEP=  2.0000     CRIT= 0.77583E-05
F=   196.82    FNEW=   196.82    ISQZ=   1 STEP=  1.0000     CRIT= 0.11809E-04
F=   196.82    FNEW=   196.82    ISQZ=   1 STEP=  1.0000     CRIT= 0.27549E-05

CONVERGENCE ACHIEVED AFTER 24 ITERATIONS

  105 FUNCTION EVALUATIONS.

                          Full Information Maximum Likelihood
                          ===================================

                              Residual Covariance Matrix

                   EQC            EQI          EQM
EQC        0.000097038
EQI         0.00015629     0.00043666
EQM        -0.00021264    -0.00016227     0.0010587

Number of observations = 24        Log likelihood = -196.823
        Schwarz B.I.C. = 222.248

                         Standard
Parameter   Estimate       Error        t-statistic   P-value
A1          .532522      .205085        2.59658      [.009]
A2          .209422      .128189        1.63370      [.102]
A3          .385103      .085915        4.48235      [.000]
A4         -.473620      .101261       -4.67721      [.000]
A5         -.360606      .110658       -3.25875      [.001]
B1         -.244347      .473062       -.516522      [.605]
B2          1.73283      .283057        6.12182      [.000]
B3         -.252597      .074006       -3.41320      [.001]
B4          .227535      .108505        2.09700      [.036]
B5         -.225663      .226928       -.994428      [.320]
C1         -2.21637      1.37875       -1.60752      [.108]
C2          4.87888      .520245        9.37804      [.000]
C3         -.091867      .154722       -.593753      [.553]
C4         -.023498      .071327       -.329438      [.742]
C5          .460320      .277728        1.65745      [.097]
C6         -.197906      .116505       -1.69869      [.089]

Standard Errors computed from quadratic form of analytic first
derivatives (Gauss)

Equation: EQC
Dependent variable: DLC

       Mean of dep. var. = .018001      Std.  error of regression = .985081E-02
  Std. dev. of dep. var. = .011672                      R-squared = .257131
Sum of squared residuals = .232892E-02             Durbin-Watson = .846172
   Variance of residuals = .970385E-04

Equation: EQI
```




```
Dependent variable: DLI

       Mean of dep. var. = .022827      Std.  error of regression = .020896
  Std. dev. of dep. var. = .039421                      R-squared = .743255
Sum of squared residuals = .010480                  Durbin-Watson = .834280
   Variance of residuals = .436658E-03

Equation: EQM
Dependent variable: DLM

       Mean of dep. var. = .046139       Std. error of regression = .032538
  Std. dev. of dep. var. = .043833                      R-squared = .664563
Sum of squared residuals = .025409                  Durbin-Watson = 1.31892
   Variance of residuals = .105871E-02
```

Références bibliographiques

[DAV 2003] R. Davidson, J.G. Mackinnon, *Estimation And Inference In Econometrics*, Oxford University Press, 2003.

[GRE 2003] W.Greene, *Econometric Analysis*, 5th Edition, Prentice Hall, 2003.

[HAM 1994] J. Hamilton, *Time Series Analysis*, Princeton University Press, 1994

[JOH 1997] J. Johnston, J. DiNardo, *Econometric Methods*, 4th ed., McGraw Hill International Editions, 1997.

[ZEL 1962] A. Zellner, An Efficient Method of Estimating Seemingly Unrelated Regressions, and Tests for Aggregation Bias, dans *Journal of the American Statistical Association*, 57, p. 348–368, 1962.

Chapitre 5

Tests de racine unitaire et modèles ARIMA

Ce chapitre définit les processus stochastiques stationnaires et non stationnaires, et insiste sur la distinction entre processus à tendance déterministe et à tendance stochastique. Il présente également les principaux tests de racine unitaire, décrit les modèles ARMA et ARIMA et explique comment les utiliser pour modéliser l'évolution d'une série temporelle.

1 Processus stochastiques

Les séries temporelles observées sont considérées comme des réalisations de processus stochastiques, définis jusqu'à présent (voir chapitre 1) comme des successions temporelles de variables aléatoires représentant un même concept. L'objet de cette section est de préciser la définition des processus stochastiques et d'examiner leurs propriétés.

Un processus stochastique X ou $\{X_t\}$ est une variable aléatoire indicée par le temps, multivariée et dont chaque composante univariée représente toujours le même concept économique quelle que soit la période considérée.

$$X = \{X_t\}_{t=1,\dots,n} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \qquad (5.1)$$

La réalisation d'un processus stochastique X est une série temporelle observée, dont chaque composante est une réalisation d'une composante différente de X :

- La variable aléatoire univariée X_1 a une réalisation x_1.
- La variable aléatoire univariée X_2 a une réalisation x_2.
- ...
- La variable aléatoire univariée X_n a une réalisation x_n.

La série temporelle observée $x = \{x_t\}_{t=1,\ldots,n} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ est donc une réalisation du processus stochastique $X = \{X_t\} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$. Toute série temporelle observée est considérée comme la réalisation unique d'un processus stochastique, qui aurait pu générer d'autres séries de réalisations.

Exemple

La série de l'investissement, telle qu'elle est publiée par l'INSEE dans les comptes nationaux, est considérée comme la réalisation d'un processus stochastique I. L'investissement observé à chaque période est considéré comme la réalisation d'une variable aléatoire univariée « investissement à la période t ». En d'autres termes, à chaque période, on a une variable aléatoire « investissement » distincte et pour chaque variable « investissement », on n'a qu'une réalisation. Ainsi l'investissement observé en 70, i_{70}, est une réalisation d'une variable aléatoire univariée I_{70}. L'investissement observé en 71, i_{71}, est une réalisation d'une variable aléatoire univariée I_{71}, et ainsi de suite.

Les variables aléatoires « investissement de 70 », I_{70}, et « investissement de 71 », I_{71}, et plus généralement les variables « investissement » de toutes les périodes, sont distinctes. Le vecteur formé par toutes ces variables aléatoires univariées est une variable aléatoire multivariée, le processus stochastique de l'investissement :

$$\{I_t\}_{t=70,\ldots,90} = \underbrace{\begin{pmatrix} I_{70} \\ I_{71} \\ \vdots \\ I_{90} \end{pmatrix}}_{\text{Processus stochastique}} \xrightarrow[\textit{réalisation}]{} \underbrace{\begin{pmatrix} i_{70} \\ i_{71} \\ \vdots \\ i_{90} \end{pmatrix}}_{\text{Série temporelle observée}} = \{i_t\}_{t=70,\ldots,90}$$

Un processus stochastique X ou $\{X_t\}$ est caractérisé par :

- une espérance $E(X_t)$ à chaque période t : $E(X_1), E(X_2)\ldots$
- une variance $V(X_t)$ à chaque période t : $V(X_1), V(X_2)\ldots$
- des covariances $\mathrm{Cov}(X_t, X_{t-\theta})$ pour toutes les périodes t et tous les retards θ : $\mathrm{Cov}(X_2, X_1), Cov(X_3, X_1)$, $\mathrm{Cov}(X_3, X_2)\ldots$



1.1 Processus stochastiques stationnaires

Un processus stochastique X ou $\{X_t\}$ est stationnaire si :

- $E(X_t) = \mu_X \quad \forall t,$
- $V(X_t) = \sigma_X^2 \quad \forall t\,;$
- $\text{Cov}(X_t, X_{t-\theta}) = \nu(\theta) \quad \forall t, \forall \theta.$

En d'autres termes, X est stationnaire si les conditions suivantes sont réunies :

- L'espérance est constante (elle est inchangée dans le temps).
- La variance est constante dans le temps.
- La covariance entre une composante d'une date t et une composante d'une autre date $t - q$ ne dépend que de l'écart de temps (le « retard ») θ entre les dates, et non de la date t elle-même : par exemple $\text{Cov}(X_{72}, X_{70}) = \text{Cov}(X_{89}, X_{87})$.

Une série temporelle observée est donc une réalisation d'un processus stochastique stationnaire si elle fluctue autour d'une valeur moyenne stable, si l'amplitude moyenne de ses fluctuations reste stable dans le temps, et si la manière dont ses valeurs sont liées aux valeurs précédentes se répète de façon stable dans le temps. Par exemple, une série observée ayant une forme sinusoïdale régulière est une réalisation d'un processus stochastique stationnaire.

La covariance $\text{Cov}(X_t, X_{t-\theta})$ entre deux composantes de dates différentes (t et $t - \theta$) d'un processus stochastique est appelée **autocovariance**. Lorsque le processus est stationnaire, cette autocovariance est la même dès lors que l'écart entre les dates est inchangé.

Le coefficient de corrélation entre deux composantes de dates différentes, appelé coefficient d'**autocorrélation**, est défini par :

$$\forall t \;:\; \rho_{X_t, X_{t-\theta}} = \frac{\text{Cov}\,(X_t, X_{t-\theta})}{\sqrt{V\,(X_t)\,V\,(X_{t-\theta})}} \tag{5.2}$$

Lorsque le processus est stationnaire, ce coefficient est dit d'**autocorrélation d'ordre $\boldsymbol{\theta}$** ; il est le même dès lors que l'écart entre les dates est inchangé :

$$\forall t \;:\; \rho_{X_t, X_{t-\theta}} = \frac{\nu\,(\theta)}{\sigma_X^2} = \rho_\theta \tag{5.3}$$

Puisque $\nu(0) = \text{Cov}(X_t, X_{t-0}) = \text{Cov}(X_t, X_t) = V(X_t) = \sigma_X^2$, on a toujours $\rho_0 = 1$.

Lorsque le processus stochastique est stationnaire, on peut dessiner le graphique des différents coefficients d'autocorrélation en fonction du retard (ou écart de dates) θ. Ce graphique est l'**autocorrélogramme théorique** du processus. Il indique comment une composante de date quelconque du processus est liée linéairement aux composantes des dates précédentes, pour des retards croissants. C'est une représentation graphique de la « mémoire » du processus, qui montre dans quelle mesure ses réalisations courantes sont influencées par ses réalisations passées.

> **Remarque**
>
> Le concept d'autocorrélogramme n'a de sens que pour un processus stationnaire. En effet, en cas de non-stationnarité, les coefficients d'autocorrélation dépendent non seulement du retard, mais également des dates elles-mêmes. La non-stationnarité implique que, pour une même valeur de θ :
>
> $$\rho_{X_t, X_{t-\theta}} \neq \rho_{X_{t+1}, X_{t+1-\theta}} \neq \rho_{X_{t+2}, X_{t+2-\theta}} \neq \cdots$$
>
> Le coefficient d'autocorrélation n'est pas identique pour toutes les dates t. Par contre, la stationnarité implique que, pour une même valeur de θ :
>
> $$\rho_{X_t, X_{t-\theta}} = \rho_{X_{t+1}, X_{t+1-\theta}} = \rho_{X_{t+2}, X_{t+2-\theta}} = \cdots = \rho_\theta$$
>
> Le concept d'autocorrélogramme a alors un sens.

Un cas particulier de processus stochastique stationnaire est le bruit blanc. Un processus stochastique X ou $\{X_t\}$ est un **bruit blanc** si :

- $E(X_t) = 0 \;\; \forall t$;
- $V(X_t) = \sigma_X^2 \;\; \forall t$;
- $\mathrm{Cov}(X_t, X_{t-\theta}) = 0 \;\; \forall t, \forall \theta \neq 0$.

Dans ce cas particulier, l'espérance des composantes est nulle à chaque période et les composantes de dates différentes ne sont pas liées linéairement. Une série observée qui correspond à une succession de valeurs de moyenne 0, de dispersion stable dans le temps et sans lien linéaire les unes avec les autres, est une réalisation d'un bruit blanc.

> **Remarque**
>
> Un processus stochastique est non stationnaire lorsque l'une des trois conditions de la stationnarité n'est pas remplie. Les cas les plus fréquents de non-stationnarité sont dus à une espérance ou à une variance qui varie dans le temps, ou aux deux à la fois.

1.2 Typologie des processus stochastiques stationnaires

Représentation de Wold et filtres linéaires

Le théorème de Wold prouve mathématiquement qu'un processus stochastique stationnaire peut toujours être généré par la somme d'une constante et d'une combinaison linéaire, parfois infinie, des composantes courante et passées d'un bruit blanc. Cela se traduit en termes mathématiques de la façon suivante :

$$\{X_t\} \quad \text{stationnaire} \quad \Rightarrow$$
$$\exists \quad \text{un bruit blanc } \{a_t\} \text{ et des coefficients } \mu_X, \psi_1, \psi_2 \ldots \text{ tels que :}$$
$$\forall t \;:\; X_t = \mu_X + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

Cette équation est la **représentation de Wold** du processus stochastique stationnaire $\{X_t\}$. Ici, l'espérance de X_t est représentée par μ_X et la variance de X_t par $\sigma_a^2 \left(\sum_{i=1}^{\infty} \psi_i^2 + 1\right)$. On distingue un processus stochastique stationnaire d'un autre par :

la spécificité des paramètres qui déterminent la fonction de densité du bruit blanc $\{a_t\}$ qui le génère ;

la spécificité des valeurs des coefficients $\mu, \psi_1, \psi_2 \ldots$ Sur cette base, on met en évidence trois types de processus stochastiques stationnaires :

- les processus autorégressifs $AR(p)$ ou $ARMA(p,0)$;
- les processus en moyenne mobile $MA(q)$ ou $ARMA(0,q)$;
- les processus mixtes (autorégressifs et en moyenne mobile) $ARMA(p,q)$.

Processus stochastiques autorégressifs stationnaires

Les processus **autorégressifs** peuvent s'écrire comme la somme d'une constante, de la valeur courante d'un bruit blanc et d'une combinaison linéaire **finie** de leurs valeurs passées.

Un processus stochastique $\{X_t\}$ est dit **autorégressif d'ordre *p***, et noté $AR(p)$ ou $ARMA(p,0)$, si :

$$\forall t\ :\ X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + a_t$$

où le processus stochastique $\{a_t\}$ est un bruit blanc.

Exemple

Processus $AR(1)$ ou $ARMA(1,0)$:

$$\forall t\ :\ X_t = \mu + \varphi_1 X_{t-1} + a_t$$

où $\{a_t\}$ est un bruit blanc.

Exemple

Processus $AR(2)$ ou $ARMA(2,0)$:

$$\forall t\ :\ X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + a_t$$

où $\{a_t\}$ est un bruit blanc.

Soit l'opérateur de retard L, défini par la relation suivante : $L^i X_t = X_{t-i}$. Par exemple, $LX_t = X_{t-1}, L^2 X_t = X_{t-2} \ldots$

Si l'on utilise cette notation dans l'équation générique précédente, on obtient :

$$\forall t\ :\ X_t = \mu + \varphi_1 LX_t + \varphi_2 L^2 X_t + \cdots + \varphi_p L^p X_t + a_t$$

c'est-à-dire :

$$\forall t\ :\ X_t - \varphi_1 LX_t - \varphi_2 L^2 X_t - \cdots - \varphi_p L^p Xt = \mu + a_t$$

ou :

$$\forall t\ :\ (1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)Xt = \mu + a_t$$

Le polynôme de degré p en l'opérateur de retard $L(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)$ est caractéristique du processus $AR(p)$.

Pour qu'un processus stochastique autorégressif soit **stationnaire**, les racines [1] du polynôme caractéristique $(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)$ doivent être en dehors du cercle unité. Quand ces racines sont réelles, elles doivent être strictement inférieures à -1 ou strictement supérieures à 1.

1. Les racines du polynôme $(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)$ sont les valeurs de L telles que $(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p) = 0$.

Exemple

Soit un processus $AR(1)$ ou $ARMA(1,0)$, $\{X_t\}$, défini par la relation :

$$X_t = \mu + \varphi_1 X_{t-1} + a_t$$

où a est un bruit blanc. $\{X_t\}$ se représente aussi de la manière suivante avec l'opérateur de retard :

$$(1 - \varphi_1 L)X_t = \mu + a_t$$

Son polynôme caractéristique est alors $(1 - \varphi_1 L)$ dont la racine[1] est $\frac{1}{\varphi_1}$. Ce processus stochastique $\{X_t\}$ est donc stationnaire si $\frac{1}{\varphi_1} < -1$ ou si $\frac{1}{\varphi_1} > 1$, et donc si $-1 < \varphi_1 < 1$.

Un processus stochastique autorégressif d'ordre p a par définition une représentation autorégressive finie (il dépend de ses valeurs passées jusqu'à un retard fini p) :

$$\forall t \ : \ X_t = \mu + \varphi_1 L X_t + \varphi_2 L^2 X_t + \cdots + \varphi_p L^p X_t + a_t \tag{5.4}$$

On montre que si ce processus est stationnaire, il a également une représentation en forme de moyenne mobile infinie des valeurs courante et passées du bruit blanc $\{a_t\}$. On dit donc qu'un processus AR stationnaire est **inversible**. Cela s'illustre par l'exemple simple d'un processus $AR(1)$:

$$\begin{aligned}
X_t &= \mu + \varphi_1 X_{t-1} + a_t \\
&= \mu + \varphi_1(\mu + \varphi_1 X_{t-2} + a_{t-1}) + a_t \\
&= \mu + \varphi_1\mu + \varphi_1^2 X_{t-2} + \varphi_1 a_{t-1} + a_t \\
&= \mu + \varphi_1\mu + \varphi_1^2(\mu + \varphi_1 X_{t-3} + a_{t-2}) + \varphi_1 a_{t-1} + a_t \\
&= \mu + \varphi_1\mu + \varphi_1^2 + \varphi_1^3 X_{t-3} + \varphi_1^2 a_{t-2} + \varphi_1 a_{t-1} + a_t \\
&= (1 + \varphi_1 + \varphi_1^2)\mu + \varphi_1^3 X_{t-3} + \varphi_1^2 a_{t-2} + \varphi_1 a_{t-1} + a_t \\
&= \ldots \\
&= \mu \sum_{i=0}^{\infty} \varphi_1^i + \lim_{i\to\infty} \varphi_1^i X_{t-i} + \sum_{i=0}^{\infty} \varphi_1^i a_{t-i} \ldots \\
&= \mu \sum_{i=0}^{\infty} \varphi_1^i + \sum_{i=0}^{\infty} \varphi_1^i a_{t-i} \quad \text{si} \quad -1 < \varphi_1 < 1 \\
&= \frac{\mu}{1 - \varphi_1} + \sum_{i=0}^{\infty} \varphi_1^i a_{t-i} \quad \text{si} \quad -1 < \varphi_1 < 1
\end{aligned}$$

Il s'agit de la forme particulière que prend la représentation de Wold d'un processus stochastique $AR(1)$ stationnaire : $\forall t \ : \ X_t = \mu_X + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$ où $\mu_X = \frac{\mu}{1 - \varphi_1}$ et $\psi_i = \varphi_1^i \forall i$.

On peut obtenir ce résultat d'une autre manière, en utilisant l'opérateur de retard. On sait en effet que $(1 - \varphi_1 L)X_t = \mu + a_t$ et donc que $X_t = \frac{1}{1 - \varphi L}(\mu + a_t)$.

1. Si $L = \frac{1}{\varphi_1}$, alors $(1 - \varphi_1 L) = \left(1 - \varphi_1 \frac{1}{\varphi_1}\right) = 0$.



L'expression $\dfrac{1}{1-\varphi L}$ peut se calculer aisément par une division polynomiale usuelle :

$$
\begin{array}{c|l}
1 & (1-\varphi_1 L) \\
\hline
\underline{(1-\varphi_1 L)} & 1+\varphi_1 L+\varphi_1^2 L^2+\cdots \\
\varphi_1 L & \\
\underline{\varphi_1 L-\varphi_1^2 L^2} & \\
\varphi_1^2 L^2 & \\
\underline{\varphi_1^2 L^2-\varphi_1^3 L^3} & \\
\varphi_1^3 L^3 & \\
\vdots &
\end{array}
$$

et donc :

$$
\begin{aligned}
X_t &= \frac{1}{1-\varphi_1 L}(\mu+a_t) = \left(1+\varphi_1 L+\varphi_1^2 L^2+\cdots\right)(\mu+a_t) \\
&= \left(1+\varphi_1 L+\varphi_1^2 L^2+\cdots\right)\mu+\left(1+\varphi_1 L+\varphi_1^2 L^2+\cdots\right)a_t \\
&= \left(1+\varphi_1+\varphi_1^2+\cdots\right)\mu+\left(a_t+\varphi_1 L a_t+\varphi_1^2 L^2 a_t+\cdots\right) \quad \text{car } \; L^i\mu=\mu \;\; \forall i \\
&= \frac{\mu}{1-\varphi_1}+\left(a_t+\varphi_1 a_{t-1}+\varphi_1^2 a_{t-2}+\cdots\right)
\end{aligned}
$$

L'autocorrélogramme simple d'un processus $AR(p)$ stationnaire décline rapidement avec l'augmentation du retard, mais il ne s'annule jamais :

$$\text{Pour tout } \theta, \;\; \rho_\theta \neq 0$$

Un processus $AR(p)$ reste donc lié linéairement à toutes ses valeurs passées, aussi éloignées soient-elles.

Pour un processus $AR(1)$ stationnaire, $\rho_\theta = \varphi_1^\theta$ pour tout θ, et l'autocorrélogramme peut montrer :

- soit une diminution exponentielle sans changement de signe ;
- soit une diminution exponentielle avec changement de signe à chaque retard.

En revanche, l'autocorrélogramme partiel s'annule après le premier retard :

$$\phi_1 \neq 0 \quad \text{et} \quad \phi_\theta = 0 \quad \text{pour tout } \theta > 1$$

Pour un processus $AR(2)$ stationnaire, l'autocorrélogramme simple peut montrer :

- soit une diminution exponentielle sans changement de signe ;
- soit une diminution exponentielle avec changement de signe à chaque retard ;
- soit une sinusoïde s'amoindrissant.

En revanche, l'autocorrélogramme partiel s'annule après le deuxième retard :

$$\phi_1 \neq 0\,, \;\; \phi_2 \neq 0 \quad \text{et} \quad \phi_\theta = 0 \quad \theta > 2$$

Processus stochastiques stationnaires en moyenne mobile

Les processus stochastiques **en moyenne mobile** peuvent s'écrire comme la somme d'une constante, de la valeur courante d'un bruit blanc et d'une combinaison linéaire finie des valeurs passées du bruit blanc.

Un processus stochastique $\{X_t\}$ est dit **en moyenne mobile d'ordre *q***, et noté $MA(q)$ ou $ARMA(0, q)$, si :

$$\forall t \ : \ X_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots + \theta_q a_{t-q} \qquad (5.5)$$

où le processus stochastique $\{a_t\}$ est un bruit blanc.

Exemple

Processus $MA(1)$ ou $ARMA(0, 1)$:

$$\forall t \ : \ X_t = \mu + a_t - \theta_1 a_{t-1}$$

où $\{a_t\}$ est un bruit blanc.

Exemple

Processus $MA(2)$ ou $ARMA(0, 2)$:

$$\forall t \ : \ X_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

où $\{a_t\}$ blanc.

La stationnarité n'impose aucune condition sur les coefficients, pourvu que tous les θ et la variance de a soient inférieurs à l'infini. Un modèle MA n'a de sens que s'il est inversible, ce qui implique qu'il peut s'écrire comme un modèle autorégressif avec une infinité de retards. Pour cela, les racines [1] du polynôme caractéristique $(1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)$ doivent être en dehors du cercle unité. Quand ces racines sont réelles, elles doivent être strictement inférieures à -1 ou strictement supérieures à 1.

L'autocorrélogramme simple d'un processus $MA(q)$ s'annule après le q^e retard :

$$\text{Pour tout } \ \theta < q + 1\,, \ \ \rho_\theta \neq 0\,, \qquad \text{et pour tout } \ \theta > q\,, \ \ \rho_\theta = 0$$

Un processus $MA(q)$ a donc une mémoire courte : il n'est lié linéairement qu'à son passé récent. Par contre, l'autocorrélogramme partiel d'un processus $MA(q)$ ne s'annule pas, même pour des retards très éloignés.

Pour un processus $MA(1)$, l'autocorrélogramme simple s'annule après le premier retard : $\rho_1 \neq 0$ et $\rho_\theta = 0$ pour tout $\theta > 1$, tandis que l'autocorrélogramme partiel peut montrer :

- soit une diminution exponentielle sans changement de signe ;
- soit une diminution exponentielle avec changement de signe à chaque retard.

Pour un processus $MA(2)$, l'autocorrélogramme simple s'annule après le premier retard : $\rho_1 \neq 0$, $\rho_2 \neq 0$ et $\rho_\theta = 0$ pour tout $\theta > 2$ tandis que l'autocorrélogramme partiel peut montrer :

- soit une diminution exponentielle sans changement de signe ;
- soit une diminution exponentielle avec changement de signe à chaque retard ;
- soit une sinusoïde s'amoindrissant.

1. Les racines du polynôme $(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)$ sont les valeurs de L telles que $(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p) = 0$.



Processus stochastiques mixtes stationnaires

Un processus stochastique $\{X_t\}$ est dit **mixte** ou **autorégressif en moyenne mobile d'ordres *p* et *q***, et noté $ARMA(p,q)$, si :

$$\forall t \ : \ X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots + \theta_q a_{t-q} \quad (5.6)$$

où le processus stochastique $\{a_t\}$ est un bruit blanc.

Exemple

Un processus $ARMA(1,1)$ évolue de la manière suivante :

$$\forall t \ : \ X_t = \mu + \varphi_1 X_{t-1} + a_t - \theta_1 a_{t-1}$$

où $\{a_t\}$ est un bruit blanc

Les conditions de stationnarité sont les mêmes que celles des processus autorégressifs.

1.3 Typologie des processus stochastiques non stationnaires

Processus stochastiques non stationnaires à tendance uniquement déterministe

Un processus stochastique non stationnaire à tendance uniquement déterministe évolue de la manière suivante :

$$X_t = g(t) + v_t \quad \text{pour tout} \quad t = 1 \dots n \quad (5.7)$$

où $g(t)$ est une fonction déterministe (mécanique, non aléatoire) du temps et où v_t est une composante aléatoire stationnaire. v est un processus stochastique stationnaire qui n'est pas nécessairement un bruit blanc : il peut être autocorrélé, lié à ses valeurs passées.

Le théorème de Wold montre qu'un processus stochastique stationnaire d'espérance nulle peut toujours être écrit comme une combinaison linéaire des valeurs courante et passées d'un bruit blanc :

$$v_t = u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \cdots \quad \text{pour tout } t$$

où u_t est un bruit blanc. Si l'on utilise la notation de l'opérateur de retard, on obtient :

$$v_t = (1 + \psi_1 L + \psi_2 L^2 + \cdots)u_t = \psi(L)u_t \quad \text{pour tout } t$$

où $\psi(L)$ est un polynôme en l'opérateur de retard L. Un processus stochastique non stationnaire à tendance déterministe uniquement évolue de la manière suivante :

$$X_t = g(t) + \psi(L)u_t \quad \text{pour tout} \quad t = 1 \dots n \quad (5.8)$$

Vu que la partie aléatoire v_t de X_t est stationnaire, la non-stationnarité de X_t est inhérente uniquement à sa partie déterministe $g(t)$.

Exemple

Soient $g(t)$ un polynôme du premier degré en t et v_t un bruit blanc (tous les $\psi_i = 0$ et donc $v_t = u_t$). L'évolution de X_t est décrite par la relation suivante :

$$X_t = \mu + \beta t + u_t \ \ \forall t = 1 \dots n$$

Le terme $\mu + \beta t$ est la partie déterministe de X_t, à une tendance déterministe linéaire. Le terme u_t est la partie aléatoire ou stochastique de X_t : elle est stationnaire.

Exemple

Soient $g(t)$ un polynôme du deuxième degré en t et v_t un bruit blanc (tous les $\psi_i = 0$ et donc $v_t = u_t$). L'évolution de X_t est décrite par la relation suivante :

$$X_t = \mu + \beta_1 t + \beta_1 t^2 + u_t \ \ \forall t = 1 \dots n$$

Le terme $\mu + \beta_1 t + \beta_1 t^2$ est la partie déterministe de X_t, à tendance déterministe quadratique. Le terme u_t est la partie aléatoire ou stochastique de X_t : elle suit un modèle $MA(1)$ stationnaire.

Exemple

Soient $g(t)$ un polynôme du premier degré en t et v_t un $MA(1)$ (les $\psi_i = 0$ pour $\psi_i > 1$ et donc $v_t = u_t + \psi_1 u_{t-1}$). L'évolution de X_t est décrite par la relation suivante :

$$X_t = \mu + \beta t + u_t + \psi_1 u_{t-1} \ \ \forall t = 1 \dots n$$

Le terme $\mu + \beta t$ est la partie déterministe de X_t, à tendance déterministe linéaire. Le terme $u_t + \psi_1 u_{t-1}$ est la partie aléatoire ou stochastique de X_t : elle est stationnaire.

Exemple

Soient $g(t)$ un polynôme du premier degré en t et v_t un modèle $AR(2)$ stationnaire (tous les ψ_i sont différents de 0, et $v_t = \dfrac{u_t}{1 - \varphi_1 L - \varphi_2 L^2}$). L'évolution de X_t est décrite par la relation suivante :

$$X_t = \mu + \beta t + v_t \quad \text{où} \ \ v_t = \varphi_1 v_{t-1} + \varphi_2 v_{t-2} + u_t \ \ \forall t = 1 \dots n$$

La fonction $\mu + \beta t$ est la partie déterministe de X_t, à tendance déterministe linéaire. Le terme $\dfrac{u_t}{1 - \varphi_1 L - \varphi_2 L^2}$ est la partie aléatoire ou stochastique de X_t : cette partie est stationnaire.

Pour simplifier la présentation des caractéristiques d'un processus à tendance déterministe linéaire, on se limite au cas le plus fréquent :

$$X_t = \mu + \beta t + v_t \quad \text{où} \ \ v_t = \psi(L) u_t \tag{5.9}$$

L'espérance de X_t varie dans le temps de manière déterministe :

$$E(X_t) = \mu + \beta t \ \ \forall t = 1 \dots n \tag{5.10}$$

puisque $E(v_t) = 0, \ \forall t = 1 \dots n$.

On peut expliciter la valeur de X_t et de son espérance à chaque période :

$$\begin{aligned} t=1: &\quad X_1 = \mu + \beta + v_1 \quad \text{et} \quad E(X_1) = \mu + \beta \\ t=2: &\quad X_2 = \mu + 2\beta + v_2 \quad \text{et} \quad E(X_2) = \mu + 2\beta \\ t=3: &\quad X_3 = \mu + 3\beta + v_3 \quad \text{et} \quad E(X_3) = \mu + 3\beta \\ &\vdots \end{aligned}$$

X_t augmente mécaniquement de 1β à chaque période. À chaque période, la partie aléatoire de X_t est v_t, qui est stationnaire. On dit que X_t est « stationnaire autour d'une tendance déterministe ». L'augmentation mécanique de l'espérance est la seule cause de la non-stationnarité de X_t, qui est de nature purement déterministe : la partie déterministe de X_t n'est pas stationnaire, contrairement à la partie aléatoire de X_t. La variance de X_t est constante dans le temps parce que v_t est stationnaire :

$$V(X_t) = V(v_t) = \sigma_v^2 \;\; \forall t = 1 \dots n \tag{5.11}$$

Sur les valeurs ultérieures du processus, l'effet d'un choc aléatoire u_t est temporaire, puisque son impact sur les valeurs successives $X_{t+1}, X_{t+2} \dots$ s'estompe progressivement. En effet, un choc u_t en t affecte X_{t+1} uniquement par la manière dont v_t est lié à v_{t+i}. Or v_t est stationnaire : le lien entre v_t et v_{t+i} s'estompe au fur et à mesure que l'écart de temps i s'accroît. Si est un bruit blanc, un choc u_t n'a d'effet que sur X_t, et non sur $X_{t+1}, X_{t+2} \dots$

Processus stochastiques non stationnaires à tendance stochastique

Les processus stochastiques non stationnaires à tendance stochastique sont dits **intégrés** ou **à racines unitaires**.

Un processus stochastique est non stationnaire à tendance stochastique quand une ou plusieurs racines unitaires figurent dans sa partie autorégressive, ce qui implique que la somme de ses coefficients autorégressifs est égale à 1.

Tout processus $\{X_t\}$ peut en effet être représenté approximativement par un modèle $AR(p)$, où p est suffisamment grand pour capter toute la dynamique de $\{X_t\}$:

$$\forall t \;:\; X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + u_t \tag{5.12a}$$

Si l'on utilise la notation de l'opérateur de retard, on obtient :

$$\forall t \;:\; (1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p) X_t = \mu + u_t \tag{5.12b}$$

où le processus stochastique $\{u_t\}$ est un bruit blanc.

X_t est à tendance stochastique si $\varphi_1 + \varphi_2 + \cdots + \varphi_p = 1$. Cela implique qu'une racine unitaire figure dans le polynôme caractéristique $1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p$ du processus.

Exemple

Soit le processus X_t tel que :

$$X_t = X_{t-1} + \mu + u_t$$

où u_t est un bruit blanc et μ est une dérive. Un tel processus est appelé « marche au hasard avec dérive » ou « marche aléatoire avec dérive ». La précédente équation est équivalente à :

$$X_t - X_{t-1} = \mu + u_t$$

$X_t - X_{t-1}$ est donc stationnaire et il suffit de différencier **une fois** X_t pour obtenir une expression stationnaire. On dit que le processus X_t est intégré d'ordre 1. Il peut être représenté par la relation :

$$(1 - L)X_t = \mu + u_t$$

La racine du polynôme caractéristique $(1 - L)$ est égale à 1. On dit donc que le processus X_t a une racine unitaire. On peut expliciter la valeur du processus et celle de sa variance à chaque période :

$$\begin{array}{lll} t = 1 : & X_1 = \mu + X_0 + u_1 & \text{et} \quad E(X_1) = \mu + X_0\,, \;\; V(X_1) = \sigma_u^2 \\ t = 2 : & X_2 = \mu + X_1 + u_2 = 2\mu + X_0 + u_1 + u_2 & \text{et} \quad E(X_2) = 2\mu + X_0\,, \;\; V(X_1) = 2\sigma_u^2 \\ t = 3 : & X_3 = \mu + X_2 + u_3 = 3\mu + X_0 + u_1 + u_2 + u_3 & \text{et} \quad E(X_3) = 3\mu + X_0\,, \;\; V(X_3) = 3\sigma_u^2 \\ & \ldots & \end{array}$$

Le processus X_t se forme par accumulation des chocs aléatoires u_t, si bien que sa variance augmente avec le temps. Il s'agit d'un cas de non-stationnarité d'origine aléatoire, stochastique : l'accumulation de chocs aléatoires est la tendance stochastique car elle fait augmenter la variance dans le temps. En toute période t, le processus vaut donc :

$$X_t = \mu t + X_0 + (u_1 + u_2 + u_3 + \cdots + u_t)$$

Cela implique que :

$$E(X_t) = \mu t + X_0 \quad \text{et} \quad V(X_t) = t\sigma_u^2$$

La somme $(u_1 + u_2 + u_3 + \cdots + u_t)$ est la tendance stochastique et la partie déterministe de X_t est $\mu t + X_0$. À la tendance stochastique s'ajoute une tendance déterministe si μ est différent de 0 ; sinon, le processus est une marche aléatoire sans dérive.

Les caractéristiques d'un processus stochastique X_t non stationnaire à tendance stochastique sont :

- L'espérance de X_t est constante en l'absence de dérive (terme constant) et varie dans le temps en cas de dérive : une tendance déterministe s'ajoute alors à la tendance stochastique.
- La variance de X_t varie dans le temps.
- L'effet d'un choc u_t est permanent : il affecte de la même manière toutes les valeurs X_{t+i} ultérieures.

2 Tests de racine unitaire

Lorsqu'on travaille avec une ou plusieurs séries économiques observées, il est important de déterminer, pour chaque série, si le processus stochastique dont elle est la réalisation est :

- stationnaire ;
- non stationnaire à tendance uniquement déterministe (ou stationnaire autour d'une tendance déterministe) ;
- non stationnaire à tendance stochastique.

L'importance de cette information est évidente dans les cas de figure suivants :

- On souhaite spécifier et estimer un modèle où l'évolution d'une série économique n'est « expliquée » que par son passé. Il faut d'abord « stationnariser » la série (établir une

transformation stationnaire de cette série) avant de chercher le meilleur modèle *ARMA* qui décrit l'évolution temporelle de cette transformation stationnaire. Or, si une série est non stationnaire, la nature de sa transformation stationnaire diffère en fonction de la tendance (déterministe uniquement ou stochastique) du processus de cette série.

Exemple

Soit un processus stochastique X non stationnaire à tendance stochastique, et tel que $X_t = X_{t-1} + u_t$, où u_t est un bruit blanc. Dans ce cas, la transformation $X_t - X_{t-1}$ est stationnaire. Le modèle *ARMA* appliqué à une telle transformation est alors appelé « modèle *ARIMA* », où le « *I* » rappelle qu'on modélise une transformation stationnaire d'une série « *I* »ntégrée.

Soit un processus stochastique X non stationnaire à tendance uniquement déterministe et tel que $X_t = a + bt + u_t$, où u_t est un bruit blanc. Dans ce cas, la transformation $X_t - bt$, ou $X_t - a - bt$, est stationnaire.

- On souhaite estimer un modèle linéaire reliant plusieurs séries économiques (par exemple une variable dépendante et une ou plusieurs variables explicatives) et tester des hypothèses sur les paramètres de cette relation. Les techniques d'inférence statistique à utiliser sont différentes en fonction de la nature des processus stochastiques de ces séries :
  - Les techniques d'inférence statistique classique (voir chapitres 2 et 3) sont valables à condition qu'aucune série de la relation n'ait une tendance stochastique.
  - Les techniques liées à la problématique de la cointégration s'imposent lorsque des séries de la relation ont une tendance stochastique.

Les tests de racine unitaire permettent de tester l'hypothèse H_0 qu'une série économique observée est la réalisation d'un processus stochastique non stationnaire à tendance stochastique, contre l'hypothèse H_1 que ce processus est (*i*) stationnaire à tendance uniquement déterministe ou (*ii*) stationnaire.

Ces tests sont basés sur l'estimation préalable d'un modèle autorégressif $AR(p)$ censé approcher le vrai comportement du processus stochastique dont la série observée est une réalisation.

2.1 Modélisation d'un processus par un modèle autorégressif $AR(p)$

Un processus stochastique X est bien représenté par un modèle $AR(p)$ général s'il existe des coefficients μ et φ_i, pour $i = 1 \ldots p$, tels que :

$$\forall t \;:\; X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \beta t + u_t \qquad (5.13)$$

où le processus stochastique $\{u_t\}$ est un bruit blanc. Ce modèle peut s'écrire de la manière suivante :

$$\begin{aligned}\forall t \;:\; X_t - X_{t-1} = {} & \mu + (\rho - 1)X_{t-1} + \lambda_1(X_{t-1} - X_{t-2}) + \lambda_2(X_{t-2} - X_{t-3}) \\ & + \cdots + \lambda_{p-1}(X_{t-p-1} - X_{t-p}) + \beta t + u_t \qquad (5.14)\end{aligned}$$

où le processus stochastique $\{u_t\}$ est un bruit blanc et où :

$$\begin{aligned}\rho &= \varphi_1 + \varphi_2 + \cdots + \varphi_p\,,\\ \lambda_1 &= \varphi_2 + \varphi_3 + \cdots + \varphi_p\,,\\ \lambda_2 &= \varphi_3 + \varphi_4 + \cdots + \varphi_p\,,\\ &\vdots\\ \lambda_{p-2} &= \varphi_{p-1} + \varphi_p\\ \lambda_{p-1} &= \varphi_p \end{aligned} \qquad (5.15)$$

Les deux formulations (5.13) et (5.14) du modèle $AR(p)$ sont équivalentes (ce sont deux manières différentes de formuler le même modèle, et non deux modèles différents). Une formulation implique l'autre et *vice-versa*. La formulation (5.14) est particulièrement pratique pour décrire les implications d'une racine unitaire.

Si le processus stochastique X a une racine unitaire, alors $\varphi_1 + \varphi_2 + \cdots + \varphi_p = 1$, ce qui implique $\rho = 1$ ou $\rho - 1 = 0$. Tester la présence d'une racine unitaire revient donc à tester l'hypothèse que le coefficient de X_{t-1} dans la formulation (5.14) du modèle $AR(p)$, en l'occurrence $\rho - 1$, est égal à 0.

Si le processus X a une racine unitaire, son évolution peut donc être représentée par :

$$\forall t\ :\ X_t - X_{t-1} = \mu + \lambda_1(X_{t-1} - X_{t-2}) + \lambda_2(X_{t-2} - X_{t-3}) + \cdots + \lambda_{p-1}(X_{t-p-1} - X_{t-p}) + u_t \qquad (5.16)$$

Cela équivaut à :

$$\begin{aligned}\forall t\ :\ &(X_t - X_{t-1}) - (X_{t-1} - X_{t-2})\\ &= \mu + (\lambda_1 + \lambda_2 + \cdots \lambda_{p-1} - 1)(X_{t-1} - X_{t-2})\\ &\quad + (\lambda_2 + \cdots + \lambda_{p-1})\big((X_{t-2} - X_{t-3}) - (X_{t-3} - X_{t-4})\big) + \cdots\\ &\quad + \lambda_{p-1}\big((X_{t-p+2} - X_{t-p+1}) - (X_{t-p+1} - X_{t-p})\big) + u_t \end{aligned} \qquad (5.17)$$

La transformation $X_t - X_{t-1}$ de X_t est stationnaire si $\lambda_1 + \lambda_2 + \cdots \lambda_{p-1} < 1$. Dans ce cas, on dit que X est intégré d'ordre 1 : il suffit de le différencier une fois pour le rendre stationnaire.

Si $\lambda_1 + \lambda_2 + \cdots + \lambda_{p-1} = 1$, la transformation $X_t - X_{t-1}$ a aussi une racine unitaire. La nouvelle transformation $(X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$ de X_t est alors stationnaire si $(\lambda_2 + \cdots + \lambda_{p-1}) + (\lambda_3 + \cdots + \lambda_{p-1}) + \cdots + \lambda_{p-1} < 1$. Dans ce cas, on dit que X est intégré d'ordre 2 : il suffit de le différencier deux fois pour le rendre stationnaire.

Remarque

L'ordre d d'intégration de X (il faut différencier d fois X pour le rendre stationnaire) ne dépend pas de p ou de $p-1$ (il faut spécifier un modèle autorégressif avec p retards ou $p-1$ différences retardées pour que le terme d'erreur soit un bruit blanc).

2.2 Tests de racine unitaire contre stationnarité ou tendance déterministe

Cette section décrit les tests de l'hypothèse de racine unitaire contre l'hypothèse de stationnarité ou l'hypothèse de tendance déterministe. Pour détecter si un processus a une tendance stochastique, c'est-à-dire s'il a au moins une racine unitaire, et donc s'il

est intégré, on utilise souvent les tests de Dickey et Fuller [FULL 1996] [DIC 1979], basés sur une estimation préalable d'un modèle autorégressif $AR(p)$ avec suffisamment de retards, donc avec une valeur suffisante de p, pour que ce modèle soit une approximation acceptable du vrai processus stochastique dont la série observée est une réalisation. Parmi les différentes catégories de tests proposés par Dickey et Fuller, on va s'intéresser à la catégorie des tests de la t-stat, les plus utilisés. Trois types de tests figurent dans cette catégorie selon que le modèle ne contient ni constante ni tendance, qu'il contient une constante mais pas de tendance, ou une constante et une tendance. Ici, on ne présente que deux de ces trois types, ceux dont l'usage est le plus fréquent. On présente ensuite d'autres tests, plus récents, qui ont souvent de meilleures propriétés sur de petits échantillons.

Tests de Dickey et Fuller avec constante et tendance

Ces tests permettent de vérifier l'hypothèse nulle qu'un processus stochastique est à tendance stochastique, avec également une tendance déterministe, contre l'hypothèse que le processus est à tendance uniquement déterministe (donc contre l'hypothèse qu'il est « stationnaire autour d'une tendance détermininiste »).

Ces tests sont basés sur l'estimation d'un modèle $AR(p)$, avec constante et tendance, adéquat pour la série à laquelle on s'intéresse, c'est-à-dire avec suffisamment de retards (donc avec une valeur suffisante de p) pour que ce modèle soit une approximation acceptable du vrai processus stochastique dont la série observée est une réalisation. Une fois la valeur convenable de p identifiée, on dispose de l'estimation par MCO du modèle $AR(p)$ avec constante et tendance, paramétré de manière à faire apparaître des différences premières :

$$X_t - X_{t-1} = \hat{\mu} + \left(\hat{\rho} - 1\right) X_{t-1} + \hat{\lambda}_1 \left(X_{t-1} - X_{t-2}\right) + \hat{\lambda}_2 \left(X_{t-2} - X_{t-3}\right) + \cdots \\ + \hat{\lambda}_{p-1} \left(X_{t-p+1} - X_{t-p}\right) + \hat{\beta} t + e_t \tag{5.18}$$

On teste alors l'hypothèse de racine unitaire $H_0 : \rho = 1$ et $\beta = 0$, contre l'hypothèse H_1 : $r < 1$ et $\beta \neq 0$, en utilisant la statistique suivante :

$$\tau_\tau = \frac{\hat{\rho} - 1}{S_{\hat{\rho}-1}} \tag{5.19}$$

Il s'agit du coefficient estimé $\hat{\rho} - 1$ de la variable X_{t-1}, divisé par l'écart type estimé $S_{\hat{\rho}-1}$ de l'estimateur $\hat{\rho} - 1$ de $\rho - 1$. Cette statistique est donc la ***t*-stat du coefficient de X_{t-1}** dans l'équation estimée précédente.

On prouve en effet que si $\rho = 1$ et $\beta = 0$, τ_τ est distribué selon une table spéciale de Dickey et Fuller [(1)], la même table, quel que soit p. Ce test se réalise du côté gauche de la distribution spéciale de Dickey et Fuller. On prend dans la table la valeur critique telle qu'une réalisation de cette distribution spéciale a 5 chances sur 100 d'être inférieure à la valeur critique, et 95 chances sur 100 d'être supérieure. Si τ_τ est inférieur à la valeur critique (donc plus grand en valeur absolue), on rejette l'hypothèse de racine unitaire H_0 : $\rho = 1$ et $\beta = 0$. Si τ_τ est supérieur à la valeur critique (donc plus petit en valeur absolue), on ne la rejette pas. Pour un grand échantillon, la valeur critique est -3.41.

1. Puisque sous l'hypothèse nulle, la variable X a une tendance stochastique, les coefficients estimés de la régression auxiliaire (5.18) ne suivent pas l'inférence statistique classique : la statistique (5.19) ne suit donc pas une distribution Student comme dans le cas stationnaire.

En fait, des techniques différentes permettent de calculer les valeurs de la table de la distribution de τ_τ, et conduisent à des résultats légèrement distincts. Ainsi Fuller [FUL 1996] et MacKinnon [MAC 1993] proposent des tables différentes de la distribution de τ_τ.

Si $p = 1$ et donc si tous les λ_i sont nuls, le test de Dickey et Fuller est appelé « test *DF* avec tendance ». Si $p > 1$, et que des différences retardées figurent donc dans l'équation de base, on parle de « test $ADF(p-1)$ avec tendance ».

Le test n'est valable que s'il est basé sur l'estimation d'une équation autorégressive avec le p convenable. Il importe donc de bien choisir p. Pour cela, on estime l'équation précédente avec un p assez grand, on la réestime ensuite successivement avec des p plus petits, et l'on s'arrête à la valeur de p qui satisfait l'un des deux critères suivants :

- Il s'agit de la plus petite valeur p pour laquelle le résidu e de l'équation n'est pas autocorrélé.
- Il s'agit de la valeur de p qui optimise un critère d'ajustement statistique du modèle : par exemple celle qui maximise le critère d'information d'Akaike [AKA 1974] [AKA 1973], le critère AIC2 [PAN 1994], le critère SBC (Schwarz Bayesian Criterion [SCH 1978]) ou le critère HQC (Hannan-Quinn Criterion), etc.

Il est indispensable d'utiliser un test *ADF* plutôt qu'un test *DF* dès que le résidu de l'équation de base du test *DF* ($X_t - X_{t-1} = \mu + (\rho - 1)\, X_{t-1} + \beta t + u_t$) est autocorrélé. En effet, cette autocorrélation induit des distorsions importantes du seuil du test *DF*. Le test $ADF(p)$ se comporte alors mieux (moins ou pas de distorsion du seuil du test) si la valeur de p est suffisante, mais perd de la puissance à mesure que p augmente. Si p est grand par rapport à l'effectif n des observations, la puissance du test $ADF(p)$ est quasi inexistante.

De manière générale, les tests DF ou ADF ont peu de puissance sur des petits échantillons (n petit) quand il s'agit de vérifier l'hypothèse $H_0 : \rho = 1$, contre l'hypothèse $H_1 : \rho < 1$, et que le ρ « vrai inconnu » est inférieur à 1 mais proche de 1.

Si l'on ne rejette pas l'hypothèse de racine unitaire pour X_t avec le test (5.19), et si ensuite on la rejette pour la variable $X_t - X_{t-1}$ (en lui appliquant par exemple le même test (5.19) [1], le résultat du test peut être interprété de la manière suivante : on ne rejette pas l'hypothèse que la première différence de X_t est stationnaire autour d'une constante non nulle (donc que X_t est non stationnaire à tendance stochastique avec également une tendance déterministe), contre l'hypothèse que X_t a une tendance déterministe uniquement.

Remarque

Le test de Dickey et Fuller avec tendance est approprié quand le graphique de la série montre clairement que celle-ci augmente ou diminue systématiquement dans le temps.

Tests de Dickey et Fuller avec constante sans tendance

Ces tests permettent de tester l'hypothèse nulle qu'un processus stochastique est à tendance purement stochastique, sans avoir également une tendance déterministe, contre

1. Le test (5.19) appliqué à $X_t - X_{t-1}$ est calculé à partir des résultats de la régression :

$$\begin{aligned}(X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = \hat{\mu} + (\hat{\rho} - 1)(X_{t-1} - X_{t-2}) + \hat{\lambda}_1\big((X_{t-1} - X_{t-2}) - (X_{t-2} - X_{t-3})\big)\\ + \hat{\lambda}_2\big((X_{t-2} - X_{t-3}) - (X_{t-3} - X_{t-4})\big) + \cdots\\ + \hat{\lambda}_{p-1}\big((X_{t-p+1} - X_{t-p}) - (X_{t-p} - X_{t-p-1})\big) + \hat{\beta} t + e_t\end{aligned}$$

l'hypothèse que le processus est stationnaire. Ce test s'utilise donc en pratique pour une série temporelle dont le niveau moyen reste stable dans le temps, qui ne présente donc pas une tendance systématique à augmenter ou à diminuer dans le temps.

De tels tests sont basés sur l'estimation d'un modèle $AR(p)$, avec constante et sans tendance, adéquat pour la série à laquelle on s'intéresse, c'est-à-dire avec suffisamment de retards, donc avec une valeur suffisante de p, pour que ce modèle soit une approximation acceptable du vrai processus stochastique dont la série observée est une réalisation. Une fois la valeur convenable de p identifiée, on dispose de l'estimation par *MCO* du modèle $AR(p)$ avec constante et tendance, paramétré de manière à faire apparaître des différences premières :

$$X_t - X_{t-1} = \hat{\mu} + \left(\hat{\rho} - 1\right) X_{t-1} + \hat{\lambda}_1 \left(X_{t-1} - X_{t-2}\right) + \hat{\lambda}_2 \left(X_{t-2} - X_{t-3}\right) + \cdots + \hat{\lambda}_{p-1} \left(X_{t-p+1} - X_{t-p}\right) + e_t \tag{5.20}$$

On teste alors l'hypothèse de racine unitaire $H_0 : \rho = 1$ et $\mu = 0$, contre l'hypothèse H_1 : $r < 1$ et $\mu \neq 0$, en utilisant la statistique suivante :

$$\tau_\mu = \frac{\hat{\rho} - 1}{S_{\hat{\rho}-1}} \tag{5.21}$$

Il s'agit du coefficient estimé $\hat{\rho} - 1$ de la variable X_{t-1}, divisé par l'écart type estimé $S_{\hat{\rho}-1}$ de l'estimateur $\hat{\rho} - 1$ de $\rho - 1$. Cette statistique est donc la ***t*-stat du coefficient de X_{t-1}** dans l'équation estimée précédente.

On prouve en effet que si $\rho = 1$ et $\mu = 0$, τ_μ est distribué selon une table spéciale de Dickey et Fuller, la même table quel que soit p. Ce test se réalise du côté gauche de la distribution spéciale de Dickey et Fuller. On prend dans la table la valeur critique telle qu'une réalisation de cette distribution spéciale a 5 chances sur 100 d'être inférieure à la valeur critique, et 95 chances sur 100 d'être supérieure. Si τ_μ est inférieur à la valeur critique (donc plus grand en valeur absolue), on rejette l'hypothèse de racine unitaire H_0 : $\rho = 1$ et $\mu = 0$. Si τ_μ est supérieur à la valeur critique (donc plus petit en valeur absolue), on ne la rejette pas. Pour un grand échantillon, la valeur critique est -3.41.

En fait, des techniques différentes permettent de calculer les valeurs de la table de la distribution de τ_μ, et conduisent à des résultats légèrement distincts. Ainsi, Fuller [FUL 1996] et MacKinnon [MAC 1993] proposent des tables différentes de la distribution de τ_μ.

Si $p = 1$ et donc si tous les λ_i sont nuls, le test de Dickey et Fuller est appelé « test *DF* avec constante et sans tendance ». Si $p > 1$, et que des différences retardées figurent donc dans l'équation de base, on parle de « test $ADF(p - 1)$ avec constante et sans tendance ».

Le test n'est valable que s'il est basé sur l'estimation d'une équation autorégressive avec le p convenable. Il importe donc de bien choisir p. Pour cela, on estime l'équation précédente avec un p assez grand, on la réestime ensuite successivement avec des p plus petits, et l'on s'arrête à la valeur p qui satisfait l'un des deux critères suivants :

- Il s'agit de la plus petite valeur p pour laquelle le résidu e de l'équation n'est pas autocorrélé.
- Il s'agit de la valeur de p qui optimise un critère d'ajustement statistique du modèle : par exemple celle qui minimise le critère d'information AIC d'Akaike [AKA 1973] [AKA 1974], ou le critère modifié AIC2. Ce critère se définit très simplement : si l'AIC est minimisé pour un modèle à j retards ($p - 1 = j$), l'AIC2 implique un modèle à $j + 2$ retards [PANT 1994]. Il semble que l'usage de l'AIC2, au lieu de l'AIC, évite des distorsions de taille pour les tests White Symetric et Dickey et Fuller.

Si l'on ne rejette pas l'hypothèse de racine unitaire pour X_t avec le test (21), et si ensuite on la rejette pour la variable $X_t - X_{t-1}$ (en lui appliquant par exemple le même test (21)), le résultat du test peut être interprété de la manière suivante : on ne rejette pas l'hypothèse que la première différence de X_t est stationnaire avec une constante nulle (donc que X_t est non stationnaire à tendance stochastique et sans tendance déterministe), contre l'hypothèse que X_t est stationnaire.

Remarque

Le test de Dickey et Fuller sans tendance est approprié quand le graphique de la série montre clairement que celle-ci n'augmente pas ou ne diminue pas systématiquement dans le temps. Il s'applique donc à une série dont la valeur moyenne est globalement stable dans le temps.

Ce test vérifie en effet l'hypothèse que X_t a une espérance constante mais une variance qui augmente, contre l'hypothèse qu'il a une espérance et une variance constantes.

Tests de Phillips et Perron

De la même manière que les tests de Dickey et Fuller, les tests de Phillips et Perron [PHI 1988] sont calculés à partir d'une régression auxiliaire. Il s'agit en fait de la même régression que celle du test DF de Dickey et Fuller (donc le cas où $p - 1 = 0$). On distingue également trois cas pour cette régression : sans constante ni tendance, avec constante mais sans tendance, et avec constante et tendance. Les deux derniers modèles sont les plus utilisés, dans les mêmes conditions que les cas correspondants de Dickey et Fuller. Leurs hypothèses nulles sont les mêmes que précédemment.

La différence avec la méthode Dickey et Fuller est la suivante : au lieu d'ajouter $p - 1$ différences retardées à la régression, Phillips et Perron proposent une correction non paramétrique de la t-stat dans la régression de la variation de la variable sur son niveau retardé (avec ou sans constante et tendance). En effet, puisqu'on n'ajoute pas de différences retardées au modèle jusqu'à ce que le terme d'erreur soit un bruit blanc, le terme d'erreur de la régression auxiliaire est susceptible d'être autocorrélé. Il est alors justifié d'estimer les variances des estimateurs par une méthode robuste à l'autocorrélation, du type de la méthode non paramétrique de Newey et West [NEW 1987] évoquée au chapitre 4.

Les tests de Phillips et Perron dépendent donc aussi d'un paramètre comparable au « $p - 1$ » des tests de Dickey et Fuller, mais celui-ci (appelé « l ») a une autre interprétation : c'est l'ordre d'autocorrélation supposé qu'on utilise dans le calcul d'une variance estimée de long terme du résidu de la régression auxiliaire de base. Cette variance estimée intervient dans le calcul des tests de Phillips et Perron. En effet, on montre que la distribution asymptotique de la t-stat (associée à X_{t-1}) obtenue à partir de la régression auxiliaire dépend d'un paramètre $\delta = \frac{1}{2}\left(\sigma^2 - \sigma_u^2\right)$ où $\sigma_u^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E\left(u_t^2\right)$ et $\sigma^2 = \lim_{n\to\infty} E\left(\frac{1}{n}\left(\sum_{t=1}^{n} u_t\right)^2\right)$ est appelé « variance de long terme ». Or, σ_u^2 peut être estimé de manière convergente par la variance empirique $\hat{\sigma}_u^2$ des résidus estimés, et un estimateur convergent de δ est donné par la formule $\hat{\delta} = \frac{1}{n}\sum_{j=1}^{l} w_{jl} \sum_{t=j+1}^{n} \hat{u}_t \hat{u}_{t-j}$ de Newey et West [NEW 1987], ce qui fournit une estimation convergente de σ^2. Dans cette formule, le paramètre l est un retard maximum (donc un paramètre de troncature des retards) imposé arbitrairement, et w_{jl} est une certaine fonction de pondération (par

exemple $1 - \frac{j}{l+1}$) pour que l'estimation résultante de σ^2 soit positive (voir le chapitre 4 à propos de l'estimation d'une matrice de variance et de covariance par les formules de Newey et West [NEW 1987]).

On ne dispose pas de critères permettant de choisir le paramètre l convenable des tests de Phillips et Perron (alors que l'on sait choisir le bon « $p - 1$ » dans les tests de Dickey et Fuller). C'est bien entendu un inconvénient. Des simulations de Monte Carlo semblent montrer que les tests de Phillips et Perron ont une fiabilité médiocre sur de petits échantillons [SCH 1989], voire très médiocre selon Davidson et MacKinnon [MAC 1993].

Tests Weighted-Symetric

Les tests Weighted-Symetric (WS) semblent surpasser les tests de Dickey et Fuller ou de Phillips et Perron sur de petits échantillons [PAN 1994]. Ils sont basés sur une régression pondérée à « double longueur ».

La variable X est d'abord régressée par *MCO* sur une constante (pour le test WS sans tendance) et/ou sur une tendance (pour le test WS avec tendance). Le résidu E de cette première régression est ensuite utilisé à la place de X dans la régression à « double longueur ».

Remarque

Dans la première moitié de cette régression, E_t est régressé sur E_{t-1} et des retards de $E_t - E_{t-1}$, les poids étant $(t-1)/n$, où n est l'effectif des observations dans l'échantillon disponible original. Dans la seconde moitié, E_t est régressé sur E_{t+1} et des valeurs futures de $E_t - E_{t+1}$, les poids étant $\left(1 - (1-t)/n\right)$.

Le problème du choix du « bon » retard maximal p, qui implique d'ajouter $p-1$ différences retardées au modèle sous-jacent, se pose de nouveau. Les mêmes critères que ceux utilisés pour les tests *ADF* peuvent être exploités ici pour déterminer la valeur optimale de $p - 1$. Il semble que l'usage de l'AIC2, au lieu de l'AIC, évite des distorsions de taille pour les tests Weighted-Symetric.

Tests du rapport de variance (variance ratio)

Soit un modèle $AR(1)$, reparamétré de manière à faire apparaître une différence à gauche :

$$X_t - X_{t-1} = \mu + (\rho - 1)\, X_{t-1} + u_t \qquad (5.22)$$

On suppose que les termes d'erreur successifs u_t sont identiquement et indépendamment distribués, avec une variance σ_u^2. Cela implique donc entre autres que le processus $\{u_t\}$ est un bruit blanc. Le cas particulier où le processus $\{X_t\}$ a une racine unitaire, qui se caractérise par $\rho = 1$, induit la trajectoire suivante pour les X_t successifs (voir chapitre 3) :

$$X_t = X_0 + \mu t + \sum_{i=1}^{t} u_i \qquad (5.23)$$

Cette trajectoire détermine par exemple que :

$$X_{t-2} = X_0 + \mu\,(t-2) + \sum_{i=1}^{t-2} u_i \qquad (5.24)$$

$$X_t - X_{t-1} = \mu + u_t \quad \text{et} \quad X_t - X_{t-2} = 2\mu + u_t + u_{t-1} \tag{5.25}$$

$$V(X_t - X_{t-1}) = \sigma_u^2 \quad \text{et} \quad V(X_t - X_{t-2}) = 2\sigma_u^{2\,(1)} \tag{5.26}$$

La présence d'une racine unitaire dans un processus $\{X_t\}$ qui est $AR(1)$ implique la propriété suivante :

$$\frac{V(X_t - X_{t-2})}{V(X_t - X_{t-1})} = 2 \tag{5.27}$$

Le principe du test du rapport de variances [CAM 1997] est de vérifier indirectement la présence d'une racine unitaire en s'assurant que la propriété (5.27) est respectée par les données.

En fait, plusieurs tests de ce type existent car, en utilisant le même raisonnement que précédemment, on montre aisément que, pour tout entier i, l'hypothèse de racine unitaire implique la propriété :

$$V(X_t - X_{t-i}) = i\sigma_u^2 \tag{5.28}$$

et donc la propriété :

$$\frac{V(X_t - X_{t-i})}{V(X_t - X_{t-1})} = i \quad \text{pour tout} \quad i = 1, 2, 3 \ldots \tag{5.29}$$

Test de Bierens (1993)

Les tests de Bierens [BIE 1993a] appelés **HOAC** (Higher Order AutoCorrelations) et **DHOAC** sont élaborés sur la base des autocorrélations de la série pour laquelle on veut tester l'hypothèse de racine unitaire. Ils dépendent de plusieurs paramètres, à savoir μ, α et δ, tels que $\mu > 0$, $\alpha > 0$, $0 < \delta < 1$, et p, qui est l'ordre d'autocorrélation maximal utilisé dans le test : $p = (1 + \alpha n^q)$, où $q = \delta\mu/(3\mu + 2)$. Ces tests sont préprogrammés dans le logiciel Easyreg.

Les tests HOAC servent à tester l'hypothèse H_0 : X a une racine unitaire, contre l'hypothèse H_1 : X est un processus stationnaire.

Les tests DHOAC servent à tester l'hypothèse H_0 : X a une racine unitaire avec dérive, contre l'hypothèse H_1 : X est stationnaire autour d'une tendance déterministe linéaire. Ils sont donc indiqués quand la série X a visiblement un profil croissant ou décroissant.

Tests de Bierens (1997)

Ces tests de Bierens [BIE 1997] sont des extensions non linéaires des tests augmentés de Dickey et Fuller, ainsi que des tests non paramétriques. Ils servent à vérifier l'hypothèse H_0 : X a une racine unitaire avec dérive, contre l'hypothèse H_1 : X est stationnaire autour d'une tendance déterministe non linéaire. Ils sont donc indiqués quand la série X a visiblement un profil croissant ou décroissant dont l'évolution ne semble pas linéaire. Ils sont préprogrammés dans le logiciel Easyreg.

1. $V(X_t - X_{t-2}) = V(2\mu + u_t + u_{t-1}) = V(u_t + u_{t-1})$ puisque 2μ n'est pas aléatoire, donc ne varie pas, et n'a pas de dispersion. Seule la partie aléatoire d'une expression peut contribuer à sa dispersion ou variance. Puisque $\{u_t\}$ est par hypothèse une succession de variables distribuées indépendamment : $V(u_t + u_{t-1}) = V(u_t) + V(u_{t-1})$ et puisque par hypothèse u_t et u_{t-1} sont distribués identiquement, ils ont la même variance : $V(u_t) + V(u_{t-1}) = \sigma_u^2 + \sigma_u^2 = 2\sigma_u^2$.

Ces tests sont basés sur une régression auxiliaire de la forme :

$$X_t - X_{t-1} = aX_{t-1} + b_1(X_{t-1} - X_{t-2}) + \cdots + b_p(X_{t-p} - X_{t-p-1}) + b_{p+1} + b_{p+2}P(t,1) + \cdots + b_{p+m+1}P(t,m) + u_t \quad (5.30)$$

où u_t est un bruit blanc, et les P des polynômes de Chebishev en t dont on a enlevé la tendance, avec $P(t, 1)$ standardisé de manière à valoir t.

2.3 Tests de l'hypothèse de stationnarité contre l'hypothèse de racine unitaire

Test de Bierens et Guo

Six tests de Bierens et Guo [BIE 1993b] sont disponibles. Les quatre premiers vérifient l'hypothèse H_0 de stationnarité, contre l'hypothèse H_1 de racine unitaire (sans dérive).

$$H_0: \quad X_t = c + u_t \quad (5.31)$$

$$H_1: \quad X_t = X_{t-1} + u_t \quad (5.32)$$

Dans ces deux équations, u_t est un processus stochastique stationnaire d'espérance nulle.

Les deux autres tests de Bierens et Guo vérifient l'hypothèse H_0 de non-stationnarité déterministe, ou de stationnarité autour d'une tendance déterministe, contre l'hypothèse H_1 de racine unitaire avec dérive.

$$H_0: \quad X_t = c + bt + u_t \quad (5.33)$$

$$H_1: \quad X_t = b + X_{t-1} + u_t \quad (5.34)$$

Dans ces deux équations, u_t est un processus stochastique stationnaire d'espérance nulle.

Test de Kwiatkowski, Phillips, Schmidt et Shin

Le premier test de Kwiatkowski, Phillips, Schmidt et Shin [KWI 1992] vérifie l'hypothèse H_0 de stationnarité, contre l'hypothèse H_1 de racine unitaire (sans dérive).

$$H_0: \quad X_t = c + u_t \quad (5.35)$$

$$H_1: \quad X_t = X_{t-1} + u_t \quad (5.36)$$

Dans ces deux équations, u_t est un processus stochastique stationnaire d'espérance nulle.

Le deuxième test [KWI 1992] vérifie l'hypothèse H_0 de non-stationnarité déterministe, ou de stationnarité autour d'une tendance déterministe, contre l'hypothèse H_1 de racine unitaire avec dérive.

$$H_0: \quad X_t = c + dt + u_t \quad (5.37)$$

$$H_1: \quad X_t = d + X_{t-1} + u_t \quad (5.38)$$

Dans ces deux équations, u_t est un processus stochastique stationnaire d'espérance nulle.

3 Méthodes de modélisation de Box et Jenkins

Les méthodes de Box et Jenkins sont des outils systématiques qui permettent :

- de déterminer le meilleur modèle de type ARMA décrivant le processus stochastique d'une série observée ou d'une transformation stationnaire de celle-ci [1] ;
- d'estimer ce modèle ;
- de l'utiliser pour extrapoler les valeurs de la série.

On recourt à cette méthode pour obtenir un modèle « expliquant » les fluctuations d'une série uniquement en fonction de son passé et réaliser ensuite des extrapolations des valeurs de la variable. Si la série laisse entrevoir un profil qui se répète de manière assez régulière, en d'autres termes si des récurrences importantes peuvent être exploitées pour expliquer la série en fonction de son passé, le choix de cette méthode a un sens.

Entre un modèle qui « explique » une série uniquement en fonction de son passé, et un autre qui l'« explique » en fonction de séries qui ont une influence sur elle (facteurs explicatifs), le second offre un « degré d'explication » supérieur. Mais on dispose rarement de données sur ces variables explicatives. Dans ce cas, il faut chercher à extrapoler le mieux possible le passé de la série. Les méthodes de Box et Jenkins permettent d'identifier et d'estimer le meilleur modèle reliant la variable à son passé. Elles sont très utilisées en marketing.

3.1 Stationnarisation de la série

Puisqu'un modèle ARMA se rapporte à un processus stationnaire, il faut d'abord définir une transformation stationnaire de la série si celle-ci ne l'est pas au départ. La manière de procéder est différente selon que la série est la réalisation d'un processus non stationnaire à tendance déterministe ou à tendance stochastique. Il est donc important, au préalable, de vérifier dans quelle situation on se trouve.

On définit aisément les transformations appropriées des séries non stationnaires. Si X_t est la série à modéliser et $\tilde{X}_t$ est la transformation stationnaire de X_t, celle-ci se définit ainsi :

$$\tilde{X}_t = X_t\,, \qquad \text{si } X_t \text{ stationnaire ;} \tag{5.39a}$$

$$\tilde{X}_t = \ln(X_t)\,, \qquad \text{si } X_t \text{ non stationnaire et } \ln(X_t) \text{ stationnaire ;} \tag{5.39b}$$

$$\tilde{X}_t = (1-L)X_t\,, \qquad \text{si } X_t \text{ non stationnaire et intégré d'ordre 1 ;} \tag{5.39c}$$

$$\tilde{X}_t = (1-L)^2X_t\,, \qquad \text{si } X_t \text{ non stationnaire et intégré d'ordre 2 ;} \tag{5.39d}$$

$$\tilde{X}_t = (1-L)\ln(X_t)\,, \qquad \text{si } X_t \text{ non stationnaire et } \ln(X_t) \text{ intégré d'ordre 1 ;} \tag{5.39e}$$

$$\tilde{X}_t = (1-L)^2\ln(X_t)\,, \qquad \text{si } X_t \text{ non stationnaire et } \ln(X_t) \text{ intégré d'ordre 2 ;} \tag{5.39f}$$

$$\tilde{X}_t = X_t - \mu - \beta t\,, \qquad \text{si } X_t \text{ non stationnaire et à tendance déterministe (linéaire) uniquement.} \tag{5.39g}$$

Bien sûr, dans ce dernier cas, on doit d'abord estimer μ et β en appliquant la méthode des moindres carrés ordinaires au modèle $X_t = \mu + \beta t + u_t$.

1. Pour une autre présentation de ces méthodes, voir Phlips, Blomme, Vanden Berghe et Dor (1987), *Analyse Chronologique*, éditions De Boeck.

Quand une série mensuelle ou trimestrielle présente des variations saisonnières, la transformation stationnaire est plus complexe. Sa forme générale est alors :

$$\tilde{X}_t = (1 - L)^d (1 - L^i)^s X_t \tag{5.40}$$

ou

$$\tilde{X}_t = (1 - L)^d (1 - L^i)^s \ln(X_t) \tag{5.41}$$

où d est l'ordre de différenciation, s l'ordre de différenciation saisonnière, et i la fréquence de la série.

La différenciation saisonnière (s non nul) sert à éliminer la fluctuation saisonnière. En effet, si le profil saisonnier d'une série X_t trimestrielle a une ampleur stable dans le temps, $(1 - L^4)X_t$ $(= X_t - X_{t-4})$ n'a plus de fluctuations saisonnières alors que c'est $(1 - L^4)^2 X_t$ $(= (X_t - X_{t-4}) - (X_{t-4} - X_{t-8})$ [1]$)$ qui n'en a plus si le profil a une ampleur croissante dans le temps.Bien entendu, de telles différences saisonnières n'ont plus de fluctuations non saisonnières mais peuvent encore être non stationnaires. C'est la différenciation simple (d non nul) qui permet ensuite d'obtenir une expression stationnaire.

Exemple

Si une série X_t de fréquence trimestrielle fluctue de façon saisonnière de manière stable ($s = 1$), et si $(1 - L^4)X_t$ est intégré d'ordre 2 (ce qui se vérifie avec les méthodes décrites dans ce chapitre), alors

$$\tilde{X}_t = (1 - L)^2 (1 - L^4)^1 X_t$$

est une transformation stationnaire de X_t.

3.2 Identification du modèle

Le principe de la méthode est de calculer et de mettre en graphique les autocorrélations empiriques [2] et les autocorrélations partielles empiriques de la transformation stationnaire appropriée (identifiée à l'aide des méthodes du chapitre 6) de la série qu'on veut modéliser. On compare ces autocorrélogrammes empiriques aux autocorrélogrammes théoriques des processus $AR(1)$, $AR(2)$, $MA(1)$, $MA(2)$, $ARMA(1, 1)$, etc. et l'on choisit le modèle le plus proche pour modéliser la série transformée. Les allures des autocorrélogrammes théoriques en question ont été décrites précédemment.

Remarque

Pour bien interpréter les coefficients d'autocorrélation empiriques, il faut tenir compte du fait qu'ils sont des estimateurs des vrais coefficients d'autocorrélation de population, et qu'ils fluctuent donc autour des vrais coefficients d'autocorrélation avec une dispersion mesurée par leurs écarts types. On peut utiliser ces écarts types pour établir des intervalles de confiance pour les vrais coefficients d'autocorrélation ou des bornes pour les coefficients d'autocorrélation empiriques, à l'intérieur desquelles on ne rejette pas l'hypothèse que les vrais coefficients correspondants sont nuls.

1. $(1 - L^4)^2 = 1^2 - L^4 - L^4 + (L^4)^2 = 1 - L^4 - L^4 + L^8)$

2. On obtient les autocorrélations empiriques, qui sont les valeurs estimées des autocorrélations théoriques ρ_θ, en appliquant sur les données observées la formule de l'estimateur r_θ de ρ_θ.

Le coefficient d'autocorrélation (simple) empirique d'ordre l est calculé par :

$$r_l = \frac{\sum_{t=l+1}^{n} \left(x_t - \bar{X}\right)\left(x_{t-l} - \bar{X}\right)}{\sum_{t=l+1}^{n} \left(x_t - \bar{X}\right)^2} \tag{5.42}$$

Pour chaque retard l, le coefficient r_l d'autocorrélation empirique d'ordre l est un estimateur du vrai coefficient d'autocorrélation de population ρ_l. Une estimation approximative de l'écart type de l'estimateur r_l est donnée par la formule $\sqrt{\frac{1}{n}\left(1 + 2\sum_{j=1}^{l-1} r_j^2\right)}$ pour $l = 1, 2 \ldots \frac{n}{3}$, qui utilise une approximation de Bartlett (1946), citée par Kendall [KEN 1983]. Cette formule est utilisée entre autres par le logiciel Microfit [PES 1997].

Pour tester l'hypothèse que les p premiers coefficients d'autocorrélation de population ρ_l d'une série temporelle sont nuls, on peut utiliser le test Q d'ordre p de Box et Pierce, $Q(p) = n\sum_{j=1}^{p} r_j^2$, qui est distribué asymptotiquement (donc approximativement pour un petit échantillon) selon une Chi-2 à 2 degrés de liberté si l'hypothèse est vraie, ou la statistique Q* d'ordre p de Ljung et Box [LJU 1978], $Q^*(p) = n\,(n+2)\sum_{j=1}^{p} \frac{r_j^2}{(n-j)}$, qui est aussi distribué asymptotiquement selon une Chi-2 à p degrés de liberté si l'hypothèse est vraie. Les deux tests sont équivalents asymptotiquement, mais celui de Ljung et Box est meilleur pour de petits échantillons [HAR 1981] [KEN 1983] [PES 1997].

Si la série originale est soumise à des fluctuations saisonnières, il est souvent nécessaire de les modéliser explicitement dans la spécification, au même titre que la différenciation saisonnière évoquée précédemment.

3.3 Estimation et validation du modèle, et extrapolations sur la base du modèle estimé

On estime les coefficients « vrais inconnus » du modèle choisi à l'aide d'une technique algorithmique pour maximiser la fonction de vraisemblance [BOX 1976]. Il faut vérifier que le modèle choisi, qui vient d'être estimé, est approprié. S'il est une description convenable du vrai processus stochastique dont la série observée est une réalisation, le terme d'erreur devrait être un bruit blanc, en d'autres termes il ne devrait pas être autocorrélé.

En cas d'autocorrélation, le modèle n'est pas approprié. Il faut alors en choisir un nouveau, l'estimer et le valider. Il est parfois nécessaire de réaliser plusieurs essais avant de trouver le « bon candidat ».

Après cette étape, on peut utiliser le modèle estimé pour élaborer des extrapolations. Si, par exemple, il est de type $AR(2)$:

$$\tilde{X}_t = \mu + \hat{\varphi}_1\tilde{X}_{t-1} + \hat{\varphi}_2\tilde{X}_{t-2} + e_t \tag{5.43}$$

et qu'on l'a estimé sur les périodes t allant de 1 à n, on peut calculer des prévisions de $\tilde{X}$ sur les périodes postérieures à n de la manière suivante :

$$
\begin{aligned}
\widehat{\tilde{X}}_{n+1} &= \mu + \hat{\varphi}_1 \tilde{X}_n + \hat{\varphi}_2 \tilde{X}_{n-1} \\
\widehat{\tilde{X}}_{n+2} &= \mu + \hat{\varphi}_1 \widehat{\tilde{X}}_{n+1} + \hat{\varphi}_2 \tilde{X}_n \\
\widehat{\tilde{X}}_{n+3} &= \mu + \hat{\varphi}_1 \widehat{\tilde{X}}_{n+2} + \hat{\varphi}_2 \widehat{\tilde{X}}_{n+1} \\
\widehat{\tilde{X}}_{n+4} &= \mu + \hat{\varphi}_1 \widehat{\tilde{X}}_{n+3} + \hat{\varphi}_2 \widehat{\tilde{X}}_{n+2} \\
&\vdots
\end{aligned}
$$

Bien entendu, les extrapolations $\widehat{\tilde{X}}$ ont une imprécision potentielle qui s'explique ainsi :

- D'une part, le modèle utilisé ne décrit pas toute l'évolution de $\tilde{X}$, donc de la variance de u_t, qui est le terme d'erreur non prévisible.
- D'autre part, on ne connaît pas les vraies valeurs des coefficients μ, φ_1 et φ_2 qu'on a remplacés par des estimateurs qui sont eux-mêmes imprécis : ils ont une variance strictement positive.

En utilisant la définition de $\tilde{X}$ en fonction de X, on peut alors transformer les extrapolations $\widehat{\tilde{X}}$ de $\tilde{X}$ en extrapolations $\widehat{X}$ de X.

Résumé

Les séries temporelles observées sont des réalisations de processus stochastiques. Un processus stochastique peut être stationnaire, sous certaines conditions, ou non stationnaire avec une tendance déterministe uniquement ou avec une tendance stochastique (on dit alors que le processus a une racine unitaire, ou qu'il est intégré). Les tests de racine unitaire vérifient l'hypothèse de racine unitaire, contre l'hypothèse de stationnarité ou de non-stationnarité à tendance déterministe uniquement. La classification des processus est importante en pratique : en effet, si l'on veut étudier une relation entre plusieurs variables (processus), le type de méthode d'inférence statistique à utiliser dépend des propriétés de ces processus. Les tests de racine unitaire permettent également de déterminer une transformation stationnaire d'une série non stationnaire. C'est cette transformation stationnaire qu'il faut utiliser lorsqu'on veut spécifier et estimer un modèle de type ARMA. Un modèle ARMA appliqué à une transformation stationnaire d'un processus intégré s'appelle un modèle ARIMA.

Problèmes et exercices

Exercice 1 Efficience d'un marché boursier

Énoncé

Vérifiez l'efficience du marché boursier américain, en utilisant les données journalières du 1/1/1998 au 20/12/2000 (775 observations) issues de l'indice boursier Standard and Poors 500 (SP500). Elles sont disponibles dans le fichier SP.xls, téléchargeable sur le site Internet afférent à ce livre. Travaillez avec TSP.

Solution

Sous certaines conditions, tester l'efficience d'un marché boursier[1] revient à tester l'hypothèse que l'indice boursier correspondant a une racine unitaire. Pour comprendre cela, il est nécessaire de connaître les principes suivants de la théorie financière :

- En finance, la valeur des actions (ce qu'un investisseur rationnel est prêt à payer pour les acheter) est égale à la somme des valeurs actualisées des bénéfices futurs anticipés de toutes les périodes futures.
- Les prix des actions changent donc d'une période à l'autre, en fonction des révisions des anticipations sur les bénéfices futurs.
- Si le marché boursier est efficient, les investisseurs utilisent immédiatement toute l'information disponible pour former leurs anticipations.

De ces principes, on déduit qu'un marché boursier est efficient quand les critères suivants sont satisfaits :

- Les prix des actions à la période t et à la période $t-1$ diffèrent uniquement parce qu'une nouvelle information, imprévisible en $t-1$, est arrivée en t, provoquant un changement des anticipations de bénéfices, et donc un changement des prix des actions.
- La variation d'un indice boursier à la période t (par rapport à son niveau en $t-1$) est imprévisible à la période $t-1$ et aux périodes précédentes. En effet, cette variation résulte d'un flux d'une information nouvelle imprévisible auparavant.
- Le logarithme d'un indice boursier X est une « marche aléatoire » ou « marche au hasard » :

$$\ln(X_t) = \ln(X_{t-1}) + u_t$$

où u_t est un bruit blanc. L'indice boursier X a ainsi une racine unitaire. Un bruit blanc n'étant pas corrélé avec son passé, il représente bien mathématiquement un flux d'une information imprévisible. Cette formulation implique que la variation (logarithmique) de l'indice boursier X à la période t est imprévisible aux dates précédentes : $\ln(X_t) - \ln(X_{t-1}) = u_t$, où u_t est imprévisible par définition puisqu'il s'agit d'un bruit blanc.

1. Le lecteur souhaitant davantage d'informations sur ces notions financières peut se procurer l'ouvrage Finance de André Farber *et al.*, paru dans la même collection, aux éditions Pearson Education France. Le lecteur peut également se reporter à l'ouvrage Gestion de portefeuille, 1997, 3e éd., de C. Broquet, R. Cobbaut, R. Gillet et A. Van Den Berg, dans la collection Comptabilité, contrôle & finance, aux éditions de Boeck. La quatrième édition de cet ouvrage sortira à l'automne 2004.



Vérifier l'efficience d'un marché boursier revient donc à vérifier que l'indice boursier (ou son logarithme) est bien représenté par un modèle autorégressif avec $p - 1 = 0$, et que ce modèle a une racine unitaire.

Il faut ensuite utiliser le test de racine unitaire de Dickey et Fuller.

Si l'on représente graphiquement l'indice boursier SP500 en fonction du temps, on observe qu'il tend à augmenter nettement au cours des périodes considérées. Cela implique que son logarithme fait de même. On va tester l'hypothèse que ce logarithme a une racine unitaire avec dérive, contre l'hypothèse qu'il est non stationnaire avec tendance déterministe uniquement.

Il faut estimer des modèles autorégressifs avec tendance pour $\ln(SP500_t)$, pour différentes valeurs de $p - 1$, allant de 0 à 10 :

$$\begin{aligned}\ln(SP500_t) - \ln(SP500_{t-1}) = \mu &+ (\rho - 1)\ln(SP500_{t-1}) \\ &+ \lambda_1\big(\ln(SP500_{t-1}) - \ln(SP500_{t-2})\big) \\ &+ \lambda_2\big(\ln(SP500_{t-2}) - \ln(SP500_{t-3})\big) \\ &+ \cdots + \lambda_{p-1}\big(\ln(SP500_{t-p+1}) - \ln(SP500_{t-p})\big) + \beta t + u_t\end{aligned}$$

Il faut ensuite sélectionner le « bon » $p - 1$. Pour le modèle correspondant à la valeur de $p - 1$ retenue, on compare τ_τ à la valeur critique à 0,5 % de la table spéciale de Dickey et Fuller, ou on regarde la probabilité critique. On détermine alors si l'on rejette ou pas l'hypothèse de racine unitaire pour $\ln(SP500_t)$. Si le bon $p - 1$ vaut 0 et si l'on ne rejette pas l'hypothèse de racine unitaire, le marché est efficient.

On commence par mettre les données dans un format Excel lisible par TSP (voir chapitres 2 et 3), en procédant de la manière suivante :

- On place les valeurs successives de l'indice boursier en première colonne avec le nom de cette variable en première ligne.
- On enregistre le fichier sous le nom SP.XLS, dans le dossier C:\ du disque dur, au format Feuille de calcul Excel (peu importe la version).

Si l'on veut réaliser le test de racine unitaire manuellement, les instructions TSP nécessaires sont les suivantes :

```
freq n;
smpl 1 775;
read(file='c :\sp.xls');
lsp=log(sp);
smpl 2 775;
dlsp=lsp-lsp(-1);
trend t;
regopt(pvprint,stars) all;
noplot;
supres vcov vcor;
smpl 12 775;
olsq dlsp c t lsp(-1);
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1);
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1) dlsp(-2);
```

```
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1)-dlsp(-3);
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1)-dlsp(-4);
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1)-dlsp(-5);
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1)-dlsp(-6);
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1)-dlsp(-7);
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1)-dlsp(-8);
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1)-dlsp(-9);
cdf(dickeyf,constant,trend) @t(3);
olsq dlsp c t lsp(-1) dlsp(-1)-dlsp(-10);
cdf(dickeyf,constant,trend) @t(3);
```

On estime donc des modèles autorégressifs pour toutes les valeurs de $p-1$ allant de 0 à 10, sur un même échantillon de données, et l'on teste à chaque fois l'hypothèse de racine unitaire. Les résultats de l'estimation du modèle où $p-1=0$ sont les suivants :

```
Mean of dependent variable = .373549E-03
Std. dev. of dependent var. = .012576
   Sum of squared residuals = .119479
      Variance of residuals = .157003E-03
   Std. error of regression = .012530
                  R-squared = .996590E-02
         Adjusted R-squared = .736397E-02
    Durbin-Watson statistic = 1.97463
                  ARCH test = 14.2224 ** [.000]
                 CuSum test = .717518 [.226]
               CuSumSq test = .106228 ** [.000]
                  Chow test = 1.63542 [.178]
     LR het. test (w/ Chow) = 2.52054 [.112]
            White het. test = 54.2136 ** [.000]
 Jarque-Bera normality test = 228.336 ** [.000]
  F-statistic (zero slopes) = 3.83020 * [.022]
   Akaike Information Crit. = -5.91745
 Schwarz Bayes. Info. Crit. = -8.73711
 Log of likelihood function = 2263.47
          Estimated   Standard
Variable  Coefficient    Error         t-statistic     P-value
C         .129683       .055094        2.35385      * [.018]
T         .608486E-05   .444691E-05    1.36833        [.168]
LSP(-1)   -.018420      .792105E-02    -2.32541     * [.019]DICKEY-FULLER(CT,ASY.,1)
Test Statistic : -2.325414, Lower tail area : .41995
```

On commence par examiner les propriétés du terme d'erreur, en utilisant les tests habituels (voir chapitre 2). Ils révèlent une hétéroscédasticité (soit un rejet de l'hypothèse d'absence d'hétéroscédasticité) de type ARCH, qui induit le rejet de l'hypothèse de normalité du terme d'erreur : un processus ARCH a en effet une distribution conditionnelle normale, mais une distribution inconditionnelle (marginale) qui ne l'est pas (voir chapitre 7). L'examen des résultats des dix autres modèles (pour $p-1$ allant de 1 à 10) montre que le problème d'hétéroscédasticité persiste quelle que soit la valeur de $p-1$.

Si l'on néglige cette propriété, on constate l'absence d'autocorrélation du terme d'erreur pour tous les modèles, même celui où $p - 1 = 0$. Ce résultat suggère qu'on pourrait se baser sur le modèle où $p - 1 = 0$, ce qui est validé, entre autres, par les résultats suivants (relatifs à l'estimation du modèle où $p - 1 = 10$) : on ne rejette l'hypothèse de nullité pour aucun coefficient des variations retardées (voir les chapitres 2 et 3 à propos de ce test).

```
Mean of dependent variable = .373549E-03
Std. dev. of dependent var. = .012576
   Sum of squared residuals = .117838
      Variance of residuals = .156908E-03
   Std. error of regression = .012526
                  R-squared = .023568
         Adjusted R-squared = .796612E-02
    Durbin-Watson statistic = 1.98625
     Durbin's h alternative = 1.19625 [.232]
                  ARCH test = 11.8410 ** [.001]
                 CuSum test = .513407 [.609]
               CuSumSq test = .100406 ** [.001]
                  Chow test = 1.28459 [.216]
     LR het. test (w/ Chow) = 4.06849 * [.044]
            White het. test = 206.126 ** [.000]
 Jarque-Bera normality test = 245.995 ** [.000]
  F-statistic (zero slopes) = 1.51058 [.115]
   Akaike Information Crit. = -5.90511
 Schwarz Bayes. Info. Crit. = -8.66406
 Log of likelihood function = 2268.75

Estimated    Standard
 Variable   Coefficient     Error       t-statistic     P-value
 C          .097357       .058620       1.66082         [.093]
 T          .325612E-05   .478638E-05   .680290         [.504]
 LSP(-1)    -.013730      .843741E-02   -1.62726        [.100]
 DLSP(-1)   .127001E-02   .037068       .034262         [.923]
 DLSP(-2)   -.020995      .036935       -.568435        [.577]
 DLSP(-3)   -.068649      .036919       -1.85943        [.060]
 DLSP(-4)   .011343       .036940       .307074         [.753]
 DLSP(-5)   .948036E-02   .036751       .257961         [.785]
 DLSP(-6)   -.038645      .036744       -1.05175        [.294]
 DLSP(-7)   -.059663      .036773       -1.62247        [.101]
 DLSP(-8)   -.013601      .036655       -.371068        [.710]
 DLSP(-9)   -.061451      .036695       -1.67463        [.090]
 DLSP(-10)  .023970       .036752       .652212         [.522]
 DICKEY-FULLER(CT,ASY.,1) Test Statistic: -1.627259, Lower tail area: .78158
```

Un test du rapport de vraisemblance (voir chapitre 3) plaide également en faveur de la possibilité évoquée précédemment : il montre que le modèle où $p - 1 = 0$ n'est pas rejeté, par rapport au modèle où $p - 1 = 10$, puisque 10.56 ($= 2(2268.75 - 2263.47)$) est inférieur à la valeur critique au seuil 0,05 d'une distribution Chi-2 à 10 degrés de liberté. La valeur de τ_τ pour le modèle où $p - 1 = 0$ est alors -2.32541. La probabilité critique du test (lower tail area) est égale à 0.41995, qui est supérieur à 0.05. On ne rejette donc pas l'hypothèse de racine unitaire pour $\ln(SP500)$. Puisque, en même temps, $p - 1 = 0$, on en conclut que le marché boursier américain est efficient.

On obtient un résumé de ces résultats en écrivant un programme qui n'utilise que la commande automatique `COINT` de TSP :

```
freq n;
smpl 1 775;
read(file='C :\sp.xls');
lsp=log(sp);
coint(df,nows,nopp,const,trend,maxlag=10) lsp;
```

Les résultats sont les suivants :

```
Dickey-Fuller (tau) test

Num lags          0            1            2            3            4
alpha       0.98158      0.98138      0.98181      0.98353      0.98312
TestStat   -2.32541     -2.32989     -2.25787     -2.03202     -2.07075
P-value     0.41995      0.41749      0.45738      0.58389      0.56243
Const       0.12974      0.13110      0.12813      0.11626      0.11914
t(Const)    2.35328      2.35761      2.28605      2.06208      2.10034
Trend   6.08486D-06  6.20155D-06  5.94515D-06  4.89839D-06  5.16161D-06
t(Trend)    1.36833      1.38119      1.31224      1.07337      1.12247
Num obs   764.00000    764.00000    764.00000    764.00000    764.00000
LogLike  2263.46599   2263.48499   2263.58110   2265.12849   2265.24554
AIC        -5.91745     -5.91488     -5.91252     -5.91395     -5.91164
Var res  0.00015700   0.00015720   0.00015737   0.00015694   0.00015710

Num lags          5            6            7            8            9
alpha       0.98276      0.98354      0.98496      0.98528      0.98684
TestStat   -2.09996     -1.99268     -1.81247     -1.76351     -1.56881
P-value     0.54612      0.60542      0.69871      0.72208      0.80441
Const       0.12157      0.11620      0.10641      0.10424     0.093473
t(Const)    2.12920      2.02272      1.84407      1.79543      1.60242
Trend   5.37508D-06  4.90823D-06  4.04932D-06  3.85687D-06  2.89660D-06
t(Trend)    1.16009      1.05188      0.86316      0.81635      0.60946
Num obs   764.00000    764.00000    764.00000    764.00000    764.00000
LogLike  2265.32271   2265.69310   2267.01837   2267.08365   2268.53446
AIC        -5.90922     -5.90757     -5.90843     -5.90598     -5.90716
Var res  0.00015727   0.00015733   0.00015699   0.00015718   0.00015679

Num lags         10        Opt:2
alpha       0.98627      0.98341
TestStat   -1.62726     -2.08405
P-value     0.78158      0.55502
Const      0.097390      0.11674
t(Const)    1.66016      2.10928
Trend   3.25612D-06  5.67700D-06
t(Trend)    0.68029      1.25352
Num obs   764.00000    772.00000
LogLike  2268.75077   2285.75739
AIC        -5.90511     -5.90870
Var res  0.00015691   0.00015798
```

Les résultats des tests de racine unitaire pour chaque valeur de $p - 1$ sont les mêmes que précédemment. TSP propose automatiquement le modèle correspondant à la valeur de



$p - 1$ qui maximise le critère AIC2, en l'occurrence $p - 1 = 2$ (qui contredit l'hypothèse d'efficience du marché boursier), alors que la méthode précédente suggère le choix de $p - 1 = 0$. Les différents critères proposés pour faire le choix de $p - 1$ donnent souvent des résultats différents.

Exercice 2 Structure par termes des taux d'intérêt

Énoncé

Le différentiel entre les taux d'intérêt à long et à court terme, en données trimestrielles, a-t-il une racine unitaire?

Pour répondre à cette question, vous disposez, dans le fichier SPREAD.xls téléchargeable sur le site afférent à ce livre, des données de ce différentiel, appelé SPREAD, pour le Royaume-Uni, du deuxième trimestre 1952 au quatrième trimestre 1988.

Selon la théorie de la structure par termes des taux d'intérêt, l'écart entre taux longs et courts dépend des anticipations de la valeur des taux courts futurs et d'une prime de risque. À prime de risque constante, les changements de ce différentiel, d'une période à l'autre, ne dépendent que des révisions des anticipations des taux courts. Si les anticipations sont rationnelles, les révisions sont imprévisibles. Le différentiel d'une période est alors égal au différentiel de la période précédente, augmenté d'un terme aléatoire qui n'est pas lié à son passé, et qui représente le flux de nouvelle information arrivée durant la période considérée. Sous l'hypothèse d'anticipations rationnelles, la théorie financière de la structure par termes des taux d'intérêt implique donc que le différentiel de taux doit avoir une racine unitaire. Vérifiez ce point en utilisant les tests de Dickey et Fuller.

Solution

Un examen de la série SPREAD sous EXCEL ou TSP suggère une moyenne relativement stable dans le temps. On réalise donc des tests DF et ADF avec constante et sans tendance, de la façon suivante :

- D'une part, on applique une procédure manuelle pour les tests *DF*, *ADF*(1) et *ADF*(2) : pour plusieurs valeurs successives de $p - 1$ allant de 0 à 2, on estime l'équation $AR(p)$ reparamétrée, on récupère τ_μ et l'on compare cette valeur à la table spéciale de Dickey et Fuller.
- D'autre part, on applique une procédure automatique appelée `COINT`, pour les tests *DF* et *ADF*(1) à *ADF*(10).

Le fichier SPREAD doit se présenter sous une forme lisible par TSP. On suppose qu'il se trouve dans le répertoire C: de l'ordinateur. Le programme d'instructions TSP est le suivant :

```
freq q;
smpl 52 :1 88 :4;
read(file='uk.wks');
smpl 52 :2 88 :4;
dspread=spread-spread(-1);
?
? DF sur l'échantillon maximal pour un DF
?
olsq dspread c spread(-1);
```

```
cdf(dickeyf,notrend) @t(2);
?
? DF sur l'échantillon maximal permettant aussi un ADF(10)
?
smpl 54 :4 88 :4;
olsq dspread c spread(-1);
cdf(dickeyf,notrend) @t(2); ?
?
ADF(1) sur l'échantillon maximal pour un ADF(1)
?
smpl 52 :3 88 :4;
olsq dspread c spread(-1) dspread(-1);
cdf(dickeyf,notrend) @t(2); ?
? ADF(1) sur l'échantillon maximal permettant aussi un ADF(10)
?
smpl 54 :4 88 :4;
olsq dspread c spread(-1) dspread(-1);
cdf(dickeyf,notrend,) @t(2); ?
?
ADF(2) sur l'échantillon maximal pour un ADF(1)
?
smpl 52 :4 88 :4;
olsq dspread c spread(-1) dspread(-1) dspread(-2);
cdf(dickeyf,notrend) @t(2); ?
?
ADF(2) sur l'échantillon maximal permettant aussi un ADF(10)
?
smpl 54 :4 88 :4;
olsq dspread c spread(-1) dspread(-1) dspread(-2);
cdf(dickeyf,notrend,) @t(2); ?
?
Procedure automatique pour DF et ADF(1) jusqu'à ADF(10) sans tendance
?
smpl 52 :1 88 :4;
coint(df,nows,const,notrend,maxlag=10) spread;
```

Quand on réalise un test$ADF(p-1)$, on perd les p premières observations. On peut donc opérer les tests $ADF(p-1)$ sur un échantillon maximal, différent pour chaque valeur de p, ou sur le même échantillon, en l'occurrence le plus grand permettant de réaliser les tests. Dans ce cas, on perd p^* observations, où p^* est la plus grande valeur de p envisagé (11 si l'on va jusqu'à un $ADF(=10)$).

Dans le programme, `SPREAD` correspond à X_t, `DSPREAD` à $X_t - X_{t-1}$, `DSPREAD(-1)` à $X_{t-1} - X_{t-2}\ldots$ La variable `SPREAD(-1)` correspond à X_{t-1}.

τ_μ est la t-stat du coefficient `SPREAD(-1)`. Après chaque estimation par MCO, TSP récupère le vecteur des t-stat des coefficients dans un vecteur `@T`. La t-stat de `SPREAD(-1)`, τ_μ, est donc le deuxième élément de `@T`, soit `@T(2)`.

L'instruction `cdf` compare τ_μ, c'est-à-dire `@T(2)`, à la table de distribution spéciale de Dickey et Fuller (`dickeyf`), valable en l'absence de tendance (`notrend`). TSP donne alors

la probabilité critique : le pourcentage de distribution qui reste à gauche de τ_μ. Si cette valeur est inférieure à 0,05, on rejette l'hypothèse de racine unitaire.

Le test DF fournit les résultats suivants si le modèle $AR(1)$ correspondant (avec $p-1=0$) est estimé sur l'échantillon permettant la comparaison avec les autres modèles :

```
                        Equation 2
            Method of estimation = Ordinary Least Squares

Dependent variable: DSPREAD
Current sample: 1954:4 to 1988:4
Number of observations: 137

       Mean of dep. var. = -.029635      LM het. test = .638187 [.424]
  Std. dev. of dep. var. = .922748      Durbin-Watson = 1.83631 [.148,.654]
Sum of squared residuals = 108.663   Jarque-Bera test = 12.4303 [.002]
   Variance of residuals = .804913    Ramsey's RESET2 = .100353 [.752]
Std. error of regression = .897170    F (zero slopes) = 8.86526 [.003]
               R-squared = .061622     Schwarz B.I.C. = -.159902
      Adjusted R-squared = .054671     Log likelihood = -178.521

              Estimated   Standard
Variable     Coefficient    Error      t-statistic   P-value
C            .173571      .102631      1.69122       [.093]
SPREAD(-1)   -.139699     .046919      -2.97746      [.003]
DICKEY-FULLER(C,ASY.,0)   Test Statistic: -2.977459, Lower tail area: .03705
```

L'instruction de la procédure automatique pour réaliser des tests DF, $ADF(1) \ldots ADF(10)$ est `coint(df,nows,const,notrend,maxlag=10)`, où `df` appelle le test de Dickey et Fuller, `nows` exclut le test White Symetric, `const` exige la présence d'une constante μ dans le modèle $AR(p)$ reparamétré, `notrend` exclut une tendance dans le modèle $AR(p)$ reparamétré, et `maxlag=10` commande d'aller jusqu'à un $AR(11)$, donc jusqu'à $p-1=10$.

En réponse à cette instruction, TSP réalise d'abord les tests DF, $ADF(1) \ldots ADF(10)$ sur le plus grand échantillon utilisable. On retrouve les résultats de la procédure manuelle pour cet échantillon commun. Dans le tableau de résultats, alpha est $\hat{\rho}$, TestStat est τ_μ, p-value est la probabilité critique issue de la comparaison de TestStat à la table spéciale de Dickey et Fuller, Const est l'estimation de la constante μ dans l'$AR(p)$ et t(Const) est la t-stat de la constante. Dans la première colonne, on retrouve les résultats du test DF réalisé manuellement dont le modèle a été estimé sur l'échantillon maximal pour un $ADF(10)$.

```
Unit root tests for SPREAD

                      Dickey-Fuller (tau) test
Num lags            0            1             2             3
alpha         0.86030      0.84711       0.84106       0.83562
TestStat     -2.97746     -3.11624      -3.10162      -3.07089
P-valAsy     0.037054     0.025375      0.026433      0.028784
Const         0.17357      0.19419       0.20318       0.21125
t(Const)      1.69122      1.84804       1.88977       1.91852
Num obs     137.00000    137.00000     137.00000     137.00000
LogLike    -178.52131   -178.08534    -177.99291    -177.92382
AIC           2.63535      2.64358       2.65683       2.67042
Var res       0.80491      0.80578       0.81074       0.81606
```

```
Num lags          4           5            6            7
alpha       0.84065     0.84728      0.82551      0.85808
TestStat   -2.84534    -2.61419     -2.89234     -2.27090
P-valAsy   0.052083    0.090101     0.046246      0.18152
Const       0.20368     0.19351      0.22757      0.17658
t(Const)    1.80183     1.66775      1.92454      1.46392
Num obs   137.00000   137.00000    137.00000    137.00000
LogLike  -177.87120  -177.78101   -176.78127   -175.07283
AIC         2.68425     2.69753      2.69754      2.68719
Var res     0.82166     0.82689      0.82122      0.80725

Num lags          8           9           10        Opt:2

alpha       0.86690     0.88677      0.89571      0.84142
TestStat   -2.06027    -1.71074     -1.53475     -3.19247
P-valAsy    0.26079     0.42559      0.51626     0.020425
Const       0.16244     0.13017      0.11450      0.20620
t(Const)    1.31487     1.03605      0.89075      1.99185
Num obs   137.00000   137.00000    137.00000    145.00000
LogLike  -174.90320  -173.97810   -173.77427   -184.32808
AIC         2.69932     2.70041      2.71203      2.59763
Var res     0.81160     0.80706      0.81110      0.76533

                   Summary of Unit root tests

            TestStat     P-value     Num.lags
Dickey-F    -3.19247    0.020425      2.00000
```

À la colonne opt, TSP sélectionne le meilleur choix de $p - 1$ par rapport au critère AIC2. Il faut donc se fier au τ_μ calculé **pour cette valeur de $p - 1$**. TSP réestime alors le modèle $AR(p)$ reparamétré, sur l'échantillon maximal pour cette valeur de $p - 1$, et donne les résultats dans cette colonne. Ici, opt:2 indique que le choix porte sur un $ADF(2)$ ($p - 1 = 2$). Dans le cadre de la procédure manuelle, on obtient les valeurs de cette colonne en estimant le modèle de l'$ADF(2)$ sur l'échantillon maximal pour un $ADF(2)$, alors que l'on obtient les valeurs de la colonne 2 en estimant le modèle de l'$ADF(2)$ sur l'échantillon maximal pour un $ADF(10)$.

Exercice 3 Taux d'intérêt et rapport de variance

Énoncé

Utilisez les tests du rapport de variance pour vérifier l'hypothèse que les taux d'intérêt à court et à long terme, en France, en données annuelles, ont une racine unitaire. Ces données se trouvent dans le fichier Excel ASSUR.XLS. La variable `TMM` est le taux d'intérêt à court terme et la variable `OBL` est le taux d'intérêt à long terme. Travaillez avec TSP.

Solution

Sur le site Internet de TSP International, téléchargez la procédure TSP `vratio`, écrite par Clint Cummins, qui réalise les tests de rapport de variance. Écrivez un petit programme TSP qui applique cette procédure aux variables `TMM` et `OBL`. Vous n'avez que les

neuf premières instructions à rédiger vous-même. Pour la suite, recopiez simplement la procédure de Clint Cummins. Le programme complet est le suivant :

```
options memory=10;
options crt;
freq a;
smpl 1963 1999;
read(file='assur.xls');
title 'variance ratio tmm';
vratio tmm 2 5;
title 'variance ratio obl';
vratio obl 2 5;
INSERER ICI LES INSTRUCTIONS DE LA PROCEDURE VRATIO
```

Les résultats sont :

```
                              variance ratio tmm
                              ==================

                 Q        VRHOM      %VRHOM       VRHET      %VRHET
Value      2.00000      0.91415     0.36064     0.49141     0.62313

                 Q        VRHOM      %VRHOM       VRHET      %VRHET
Value      3.00000      0.22843     0.81931     0.14750     0.88273

                 Q        VRHOM      %VRHOM       VRHET      %VRHET
Value      4.00000     0.065000     0.94817    0.049636     0.96041

                 Q        VRHOM      %VRHOM       VRHET      %VRHET
Value      5.00000    -0.089763     0.92847   -0.076080     0.93935

                               variance ratio obl
                               ==================

                 Q        VRHOM      %VRHOM       VRHET      %VRHET
Value      2.00000      1.94849    0.051356     1.13000     0.25848

                 Q        VRHOM      %VRHOM       VRHET      %VRHET
Value      3.00000      1.96573    0.049330     1.47526     0.14014

                 Q        VRHOM      %VRHOM       VRHET      %VRHET
Value      4.00000      2.01050    0.044379      1.68581   0.091833

                 Q        VRHOM      %VRHOM       VRHET      %VRHET
Value      5.00000      1.74127    0.081636     1.61398     0.10653
```

Pour chaque valeur de i (appelée ici `Q`) allant de 2 à 5, `%VRHOM` est la probabilité critique du test `VRHOM` et `%VRHET` est la probabilité critique du test `VRHET`. Pour `TMM` et pour `OBL`, ces valeurs critiques sont toujours supérieures à 0,05. On ne rejette donc pas l'hypothèse de racine unitaire pour ces deux variables.

EXERCICE 4 TAUX D'INTÉRÊT RÉEL

Énoncé

Testez l'hypothèse que le taux d'intérêt réel à long terme en France a une racine unitaire. Cette variable, notée `OBLR`, se trouve dans le fichier ASSUR.xls, téléchargeable à partir du site afférent à ce livre. Appliquez plusieurs tests de racine unitaire préprogrammés dans Easyreg pour savoir si le taux d'intérêt réel a une tendance stochastique. Les données sont annuelles et vont de 1963 à 1999.

Solution

Préparation et lecture des données

Ouvrez le fichier ASSUR.xls sous Excel, cliquez sur Fichier puis sur Enregistrer sous, et sélectionnez CSV (séparateur point-virgule) comme type de fichier. Laissez Disque local C:\ dans la zone Enregistrer dans et conservez ASSUR comme nom de fichier. Cliquez sur Enregistrer.

Le logiciel crée alors un fichier ASSUR.csv dans C:\. Ouvrez Easyreg, puis cliquez successivement sur File/Choose an input file/Choose an Excel file in CSV format. Dans la fenêtre Guided tour qui apparaît, cliquez sur la réponse de votre choix ; si vous demandez le Guided tour, réduisez la fenêtre Guided tour on importing Excel files in CSV format pour pouvoir continuer à travailler avec Easyreg (et voir la fenêtre About CSV files). Dans la fenêtre About CSV files, cliquez sur Continue.

Vous accédez à une fenêtre Find a CSV file : sélectionnez Manually dans Search options, C: dans Choose drive, et C:\ dans Choose folder/directory. Cochez Double-click to choose file and continue. Double-cliquez sur C:\ ASSUR.csv dans Files found. Vous accédez à la fenêtre Data type of the CSV file. Cochez Annual time series, puisque les données sont annuelles. Inscrivez 1963 dans Year of first observation et cliquez sur OK. Cliquez sur Confirm, puis sur Continue. Vous accédez à Analyzing the CSV file.

Cliquez sur Code OK. La fenêtre Data written to files s'affiche. Cliquez sur Continue. Check the data apparaît. Cliquez sur Continue. La fenêtre About the start-up folder EASYREG.DAT apparaît. Cliquez sur Continue. Vous accédez à Choose drive and folder to start-up EasyReg : sélectionnez C:\, inscrivez ASSUR dans Make a new sub-folder in the Current folder Enter the folder name et cliquez sur OK. Ensuite cliquez sur Start Easyreg in current folder.

Tests de racine unitaire

Le graphique de la série OBLR en fonction du temps montre que le niveau moyen du taux d'intérêt réel est plus élevé durant les années postérieures à 1980. Cela ne signifie pas nécessairement que la série ait une véritable tendance croissante continue. D'un point de vue théorique d'ailleurs, il semble illogique qu'un taux d'intérêt réel ait tendance à augmenter indéfiniment à long terme. Pour ces raisons, il est difficile d'établir *a priori* s'il faut se fier à des tests de racine unitaire avec constante et tendance (comme le suggère le graphique) ou à des tests avec constante et sans tendance (ce que suggère le raisonnement économique). Vous allez donc faire les deux.

Pour accéder au menu des tests de racine unitaire, cliquez successivement sur Menu/Data analysis/Unit root tests (root1). La fenêtre Select time series to be tested for unit root s'affiche. Double-cliquez sur OBLR et cliquez sur Selection OK. Dans la fenêtre qui apparaît, répondez NO à la question « Do you want to use a subset of observations? ».



Cliquez sur Continue. Unit root and stationarity tests s'affiche. Cliquez sur ADF pour réaliser des tests de Dickey et Fuller. La fenêtre Augmented Dickey Fuller ADF unit root tests apparaît. Cliquez sur ADF2 pour réaliser le test sans tendance (τ_μ). Le logiciel affiche :

```
Augmented Dickey-Fuller (ADF) test 2.
Auxiliary model:
z(t)-z(t-1) = a.z(t-1) + b(1).(z(t-1)-z(t-2)) + ...
              + b(p).(z(t-p)-z(t-p-1)) + b(p+1) + u(t),
t = p+2,...,n, where u(t) is white noise.
Notice: b(p+1) is the intercept.
Null hypothesis~H(0):
z(t) is a unit root process: a = 0.
Alternative hypothesis (H1):
z(t) is stationary process: a < 0.
The test statistic is the t-value of a.
The default lag width is p = [cn^r], where:
c = 5 and r = .25.
References:
  Fuller, W.A. (1996) : Introduction to Statistical Time Series
(2nd Ed.). New York: John Wiley
  Said, S.E. and D.A.Dickey (1984): Testing for Unit Roots in
Autoregresive Moving Average of Unknown Order. Biometrika 71, 599-607
Said, S.E. (1991): Unit Root Test for Time Series Data with a Linear
Time Trend. Journal of Econometrics 47, 285-303
Warning:
This test may have low power against linear trend stationarity!
```

Puisque vous disposez de trente-sept observations, le logiciel propose un test *ADF*(p) où $p = (0{,}5)37^{0{,}25} = 12$. Cliquez sur Use this p pour choisir cette valeur de p. Attention : ce que Easyreg appelle p correspond au $p - 1$ de la partie théorique de ce chapitre et des exercices précédents. Cliquez ensuite sur Conduct test. Le logiciel affiche un long texte présentant d'abord les éléments fournis précédemment, puis la partie suivante :

```
p = 12 = [c.n^r], where c=5, r=.25, n=37
Variable to be tested:
z(t) = oblr
H0: Unit root with drift; H1: Linear trend stationarity
ADF model for z(t)-z(t-1):

          OLS estimate      t-value Asymptotic critical regions:
z(t-1)         -0.3965      -2.5426 < -2.89 (5%)
                            < -2.58 (10%)
                            p-value = 0.11000
z(t-1)-z(t-2)    -0.2449      -1.0032
z(t-2)-z(t-3)    -0.0569      -0.2915
z(t-3)-z(t-4)    -0.0708      -0.4444
z(t-4)-z(t-5)     0.0831       0.4729
z(t-5)-z(t-6)     0.4655       2.7204
z(t-6)-z(t-7)     0.4941       2.4726
z(t-7)-z(t-8)     0.1756       0.8725
z(t-8)-z(t-9)     0.1520       0.7863
z(t-9)-z(t-10)    0.1100       0.5548
```

```
z(t-10)-z(t-11)  -0.1306      -0.7205
z(t-11)-z(t-12)  -0.1415      -0.7119
z(t-12)-z(t-13)  -0.1494      -0.7294
1                 1.8743       3.2313
Residual s.e.:   88.79976E-002
R-square:         0.69849
n:               24


Test result:
HO is not rejected at the 10% significance level

Wald test that the lag width can be reduced from 12 to q:
 q   Chi-square test  d.f.  5% crit. value  10% crit. value  p-value
 11     0.532          1       3.841           2.705       0.46575
 10     0.866          2       5.991           4.605       0.64840
  9     1.148          3       7.815           6.251       0.76558
  8     2.442          4       9.488           7.780       0.65508
  7     3.833          5      11.071           9.237       0.57364
  6     4.175          6      12.591          10.645       0.65294
  5     8.115          7      14.067          12.017       0.32254
  4    14.467          8      15.507          13.361       0.07037 (*)
  3    14.641          9      16.919          14.683       0.10128
  2    14.921         10      18.307          15.987       0.13496
  1    15.410         11      19.675          17.275       0.16447
  0    15.480         12      21.026          18.549       0.21622
(*) -> significant at the 10% level


Selection of p under the null hypothesis by
the Akaike (AC), Hannan-Quinn (HQ), and Schwarz (SC)
information criteria
  p       AC        HQ        SC
  1     0.45644   0.47178   0.60246
  2     0.56751   0.59813   0.86472
  3     0.65526   0.70103   1.10917
  4     0.78883   0.84957   1.40527
  5     0.82065   0.89604   1.60581
  6     0.92069   1.01035   1.88117
  7     1.05599   1.15936   2.19883
  8     1.21533   1.33169   2.54802
  9     1.39044   1.51888   2.92100
 10     1.38759   1.52693   3.12459
 11     1.11998   1.26873   3.07259
 12     1.34935   1.50561   3.52740
Optimal p:  1         1         1
WARNING: Number of parameters in the ADF model exceeds half the number
of observations.
```

Ces résultats donnent un test *ADF*(12). $\rho - 1 = -0.3965$ et $\tau_\mu = -2.5426$. Cette valeur est supérieure à la valeur critique, qui est -2.99. L'hypothèse $\rho = 1$ n'est donc pas rejetée, ce que confirme la probabilité critique de ce test, qui est de 0.11000, soit une valeur supérieure à 0.05. La question qui se pose est la suivante : faut-il se fier à un *ADF*(12)

ou bien choisir un autre retard p dans l'équation de base ? Le deuxième tableau (Wald test...) permet de répondre à cette question. Quand q vaut 11, la ligne correspond à un test de l'hypothèse que le coefficient de la douzième variation retardée est nul (et donc qu'un $AR(11)$ suffit). Quand q vaut 10, la ligne correspond à un test de l'hypothèse que les coefficients des onzième et douzième variations retardées sont nuls (et donc qu'un $ADF(10)$ suffit), et ainsi de suite. Quand q vaut 0, le test porte sur l'hypothèse que les coefficients de la première à la douzième variations retardées sont tous nuls. Tous ces tests, sauf un, sont inférieurs à leurs valeurs critiques au seuil de 10 % de 5 %, puisque leurs probabilités critiques sont supérieures à 0,05. Leurs hypothèses nulles respectives ne sont donc pas rejetées, en particulier l'hypothèse qu'il est possible de réduire $p = 12$ à $p = 0$, c'est-à-dire qu'un test *DF* suffit[1]. Il faut donc refaire le test en choisissant $p = 0$. Pour cela, cliquez sur Done. La fenêtre Simulation of the actual p value of the *ADF* test 2 s'affiche : cliquez sur NO. Unit root and stationarity tests apparaît : cliquez sur *ADF*. Dans la fenêtre suivante, cliquez sur *ADF*2. Dans la fenêtre suivante, cliquez sur Choose another p. Ensuite, dans Enter p, saisissez 0 et cliquez sur Conduct test. Les résultats sont les suivants :

```
p =0
Variable to be tested:
z(t) = oblr
H0: Unit root with drift; H1: Linear trend stationarity
ADF model for z(t)-z(t-1):

            OLS estimate   t-value Asymptotic critical regions:
z(t-1)          -0.1428     -1.7236 < -2.89 (5%)
                            < -2.58 (10%)
                            p-value = 0.41000
1                0.5918     1.6661
Residual s.e.:  12.07081E-001
R-square:        0.08036
n:              36


Test result:
H0 is not rejected at the 10% significance level
```

L'hypothèse de racine unitaire pour le taux d'intérêt réel OBLR n'est pas rejetée. Cliquez sur Done, puis sur NO dans la fenêtre suivante. Unit root and stationarity tests apparaît. Cliquez sur ADF, puis sur *ADF*3 pour réaliser un test de Dickey et Fuller avec constante et tendance. Dans la fenêtre suivante, cliquez sur Choose another p. Ensuite, dans la fenêtre Enter p, saisissez 0 et cliquez sur Conduct test. Les résultats sont les suivants :

```
Augmented Dickey-Fuller (ADF) test 3
Auxiliary model:
z(t)-z(t-1) = a.z(t-1) + b(1).(z(t-1)-z(t-2)) + ...
 + b(p).(z(t-p)-z(t-p-1)) + b(p+1) + b(p+2).t + u(t),
t = p+2,...,n, where u(t) is white noise.
Null hypothesis H(0):
z(t) is a unit root with drift process: a = 0.
```

1. Le troisième tableau compare quelques critères usuels d'ajustement statistique pour les valeurs de p allant de 0 à 12. Ici c'est plutôt $p = 1$, donc un *ADF*(1), qui donne les meilleurs critères.

```
Alternative hypothesis H1:
z(t) is a trend stationary process: a < 0.
The test statistic is the t-value of a.
The default lag width is p = [cn^r], where
c = 5 and r = .25.
References:
Fuller, W.A. (1996): Introduction to Statistical Time Series
 (2nd Ed.). New York : John Wiley
Said, S.E. and D.A.Dickey (1984): Testing for Unit Roots in Auto-
 regresive Moving Average of Unknown Order. Biometrika 71, 599-607
Said, S.E. (1991): Unit Root Test for Time Series Data with a
 Linear Time Trend. Journal of Econometrics 47, 285-303

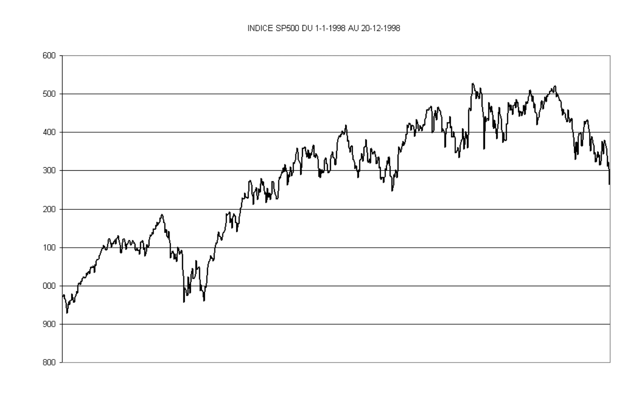
Time series to be tested for a unit root:
oblr
Available observations: t = 1(=1963),...,37(=1999)
=Chosen
p =0
Variable to be tested: z(t) = oblr
H0: Unit root with drift; H1: Linear trend stationarity
ADF model for z(t)-z(t-1):
z(t-1)      OLS estimate   t-value Critical regions: (*)
                -0.1869     -1.7677 < -3.6 (5%)
                            < -3.24 (10%)
                            p-value = 0.72000
1                0.4198     0.9579
t                0.0168     0.6803
Residual s.e.:  12.16732E-001
R-square:        0.09307
n:              36

(*) See Fuller (1996), Table 10.A.2., p.642

Test result:
H0 is not rejected at the 10% significance level
```

L'hypothèse de racine unitaire pour le taux d'intérêt réel OBLR n'est pas rejetée. Comme la *t*-stat de la tendance *t* est très faible, son coefficient n'est pas significatif. Conclusion : c'est plutôt le test ADF 2 qui est pertinent (voir figure 5.1).

Figure 5.1



Cela dit, les tests ADF 2 et ADF 3 donnent le même résultat.

Maintenant, vous allez réaliser un test HOAC de Bierens. Cliquez sur Done puis sur No. Unit root and stationarity tests s'affiche : cliquez sur Bierens HOAC. Ensuite, à la question « do you want to change Mu? », répondez No. Cliquez sur Continue. Les résultats sont les suivants :

```
Bierens' unit root tests on the basis of higher-order sample auto-
correlations: The following two tests will be conducted:
HOAC(1,1) and HOAC(2,2), which test the unit root hypothesis
against the stationarity hypothesis.
These tests depend on parameters Mu > 0, Alpha > 0, and 0 < Delta < 1,
and the lag length p of the sample autocorrelations is:
p = 1 + [Alpha * (n {\char 94} q)], where q = Delta * Mu / (3 * Mu + 2).
The default values of these test parameters are:
Mu = 2, Alpha = 5, Delta = .5
Reference:
Bierens, H.J. (1993): Higher Order Autocorelations and the Unit
Root Hypothesis. Journal of Econometrics 57, 137--160
Time series to be tested for a unit root:
oblr
Available observations : t = 1(=1963),...,37(=1999)
= Chosen
Default lag length:
p = 1 + [Alpha*(n{\char 94}q)] = 9, where q = Delta*Mu / (3*Mu + 2) = .125,
with n = 37, Mu = 2, Alpha = 5, Delta = .5
The default values for Mu, Alpha and Delta have been adopted.
Test statistics: 5\,\% Crit. Value 10\,\% Crit.  Value
HOAC(1,1) = -6.01 -14.00 -11.20
HOAC(2,2) = -8.15 -15.70 -13.10
Conclusion: Unit root hypothesis is not rejected.
```

De nouveau, l'hypothèse de racine unitaire n'est pas rejetée.

Exercice 5 Modélisation du taux de rendement des OAT

Énoncé

Spécifiez et estimez un modèle de type ARIMA pour le taux de rendement des OAT à 10 ans, en données journalières du 1/1/1998 au 22/9/2000. Utilisez la série OBL du fichier INR.xls téléchargeable à partir du site afférent à cet ouvrage.

Solution

Un graphique montrant l'évolution du taux de rendement OBL des OAT (Obligations Assimilables du Trésor) à 10 ans suggère que cette série a une moyenne en augmentation légère, ainsi qu'une variance (amplitude autour de la moyenne) en augmentation. Ces caractéristiques sont cependant peu perceptibles, si bien qu'avant de décider que cette série n'est pas stationnaire et de lui chercher une transformation stationnaire, on peut supposer qu'elle est stationnaire et calculer ses autocorrélations simples et partielles pour identifier un modèle ARMA qui lui conviendrait. On utilise en ce sens le programme TSP suivant :

```
freq n;
smpl 1 712;
read(file='inr.xls');
bjident obl;
```

Après la lecture des données de la série OBL dans le fichier INR.XLS, l'instruction BJIDENT calcule les autocorrélations simples et partielles. Les résultats sont les suivants :

```
        PROGRAM
LINE ******************************************************************
|     1  freq n;
|     2  smpl 1 712;
|     3  read(file='inr.xls');
|     4  bjident obl;
        EXECUTION

Current sample: 1 to 712

Box-Jenkins procedures
Procedure   BJIDENT

                     OPTIONS FOR THIS ROUTINE

IAC       = FALSE       NDIFF    = 0            NLAG      = 20
NLAGP     = 10          NSDIFF   = 0            NSPAN     = 0
PLOT      = TRUE        PLOTAC   = TRUE         PLTRAW    = FALSE

                          Autocorrelations

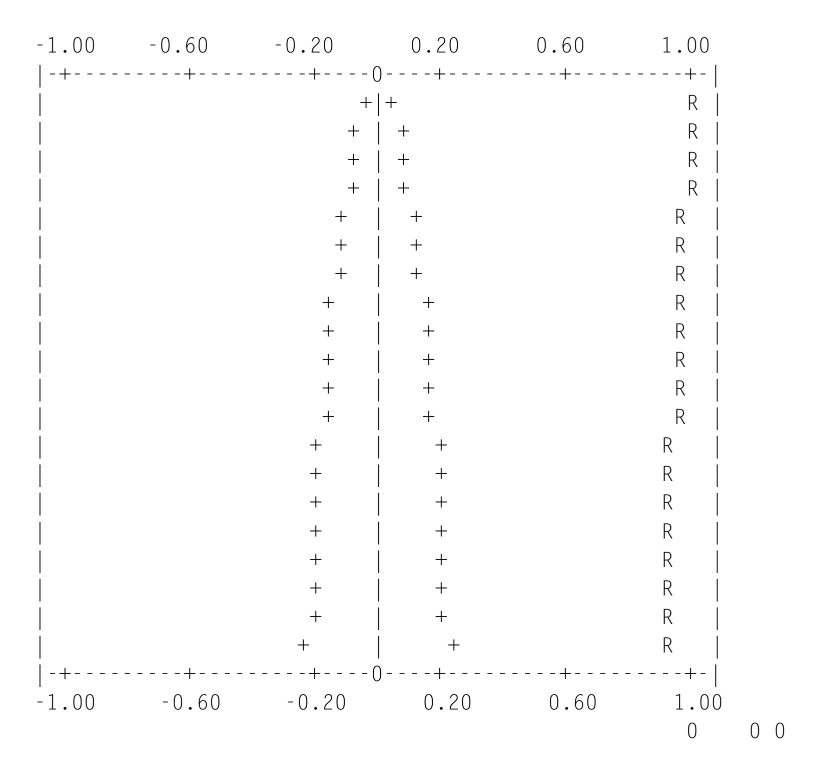
                                          0    0 0
                              Series (1-B) (1-B )  OBL
                                         Mean =    4.8555614
                               Standard Error =   0.57344979

                 | Lags  |
Autocorrelations |       |0.994    |0.988    |0.984    |0.979    |0.975
Standard Errors  | 1- 5  |0.375E-01|0.646E-01|0.832E-01|0.982E-01|0.111
Q-statistics     |       |706.     |0.141E+04|0.210E+04|0.279E+04|0.347E+04
Autocorrelations |       |0.972    |0.968    |0.964    |0.959    |0.955
Standard Errors  | 6-10  |0.122    |0.133    |0.142    |0.151    |0.160
Q-statistics     |       |0.415E+04|0.483E+04|0.550E+04|0.616E+04|0.682E+04
Autocorrelations |       |0.951    |0.947    |0.942    |0.939    |0.936
Standard Errors  | 11-15 |0.167    |0.175    |0.182    |0.189    |0.195
Q-statistics     |       |0.748E+04|0.813E+04|0.878E+04|0.942E+04|0.101E+05
Autocorrelations |       |0.932    |0.928    |0.924    |0.920    |0.915
Standard Errors  | 16-20 |0.201    |0.207    |0.213    |0.219    |0.224
Q-statistics     |       |0.107E+05|0.113E+05|0.119E+05|0.126E+05|0.132E+05

                          Partial Autocorrelations
                          ========================

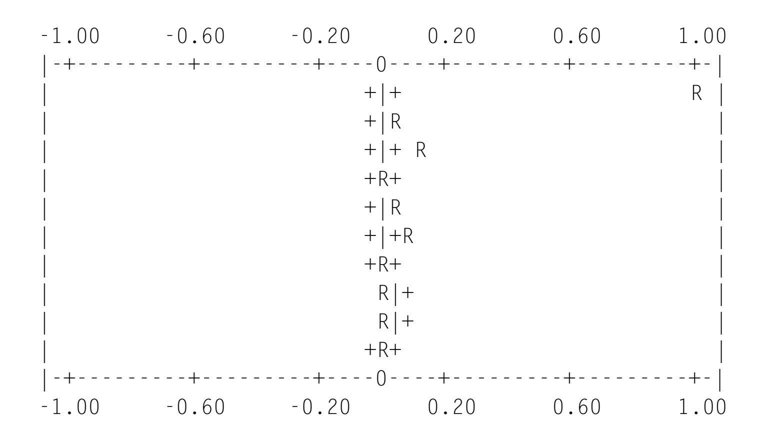
                                          0    0 0
                              Series (1-B) (1-B )  OBL
             Standard Error of Autocorrelations =  0.37476584E-01

                  |Lags |
Partial Autocorrs |1- 5 |0.994    |0.378E-01 |0.103     |0.135E-02 |0.379E-01
Partial Autocorrs |6-10 |0.620E-01|-0.177E-01|-0.253E-01|-0.223E-01|-0.775E-02
Partial Autocorrs |11-15|0.375E-01|4.86      |0.573     |5.33      |5.34
Partial Autocorrs |16-20|5.20     |5.19      |5.17      |5.14      |5.10
Partial Autocorrs |21-10|         |          |          |          |
                                              0    0 0
```



```
  Autocorrelation Function of the series (1-B) (1-B )  OBL

-1.00    -0.60     -0.20      0.20      0.60      1.00
|-+---------+---------+----0----+---------+---------+-|
|                         +|+                       R |
|                        + | +                      R |
|                        + | +                      R |
|                        + | +                      R |
|                       +  |  +                    R  |
|                       +  |  +                    R  |
|                       +  |  +                    R  |
|                      +   |   +                   R  |
|                      +   |   +                   R  |
|                      +   |   +                   R  |
|                      +   |   +                   R  |
|                      +   |   +                   R  |
|                     +    |    +                 R   |
|                     +    |    +                 R   |
|                     +    |    +                 R   |
|                     +    |    +                 R   |
|                     +    |    +                 R   |
|                     +    |    +                 R   |
|                     +    |    +                 R   |
|                    +     |     +                R   |
|-+---------+---------+----0----+---------+---------+-|
-1.00     -0.60     -0.20      0.20      0.60      1.00
                                                    0    0 0
```



```
Partial Autocorrelation Function of the series (1-B) (1-B ) OBL

-1.00     -0.60     -0.20      0.20      0.60      1.00
|-+---------+---------+----0----+---------+---------+-|
|                         +|+                       R |
|                         +|R                         |
|                         +|+ R                       |
|                         +R+                         |
|                         +|R                         |
|                         +|+R                        |
|                         +R+                         |
|                          R|+                        |
|                          R|+                        |
|                         +R+                         |
|-+---------+---------+----0----+---------+---------+-|
-1.00     -0.60     -0.20      0.20      0.60      1.00
```

Les autocorrélations simples ne diminuent pas rapidement avec l'augmentation du retard θ, mais restent élevées pour des retards θ importants : c'est une caractéristique des processus non stationnaires. Il ne faut donc pas modéliser la série OBL telle qu'elle est, mais lui trouver d'abord une transformation stationnaire.

Le deuxième graphique montre l'évolution de la première différence de OBL. Cette nouvelle série est DOBL. Cette fois, la moyenne de la nouvelle série semble bien constante dans le temps, et la variance aussi. On suppose donc qu'elle est stationnaire et l'on peut calculer ses autocorrélations simples et partielles pour identifier un modèle ARMA convenable. On utilise en ce sens le programme TSP suivant :

```
freq n;
smpl 1 712;
read(file='inr.xls');
bjident obl;
smpl 2 712;
dobl=obl-obl(-1);
bjident dobl;
```

Après la lecture des données de la série OBL dans le fichier INR.XLS, et la modification de l'échantillon pour le calcul de la première différence DOBL, l'instruction `BJIDENT` calcule les autocorrélations simples et partielles de DOBL. Les résultats sont les suivants :

```
        PROGRAM
LINE *******************************************************************
|     1  freq n;
|     2  smpl 1 712;
|     3  read(file='inr.xls');
|     4  smpl 2 712;
|     5  dobl=obl-obl(-1);
|     6  bjident dobl;
        EXECUTION

Current sample: 1 to 712
Current sample: 2 to 712

Box-Jenkins procedures
Procedure   BJIDENT

                      OPTIONS FOR THIS ROUTINE

IAC     = FALSE        NDIFF    = 0          NLAG     = 20
NLAGP   = 10           NSDIFF   = 0          NSPAN    = 0
PLOT    = TRUE         PLOTAC   = TRUE       PLTRAW   = FALSE

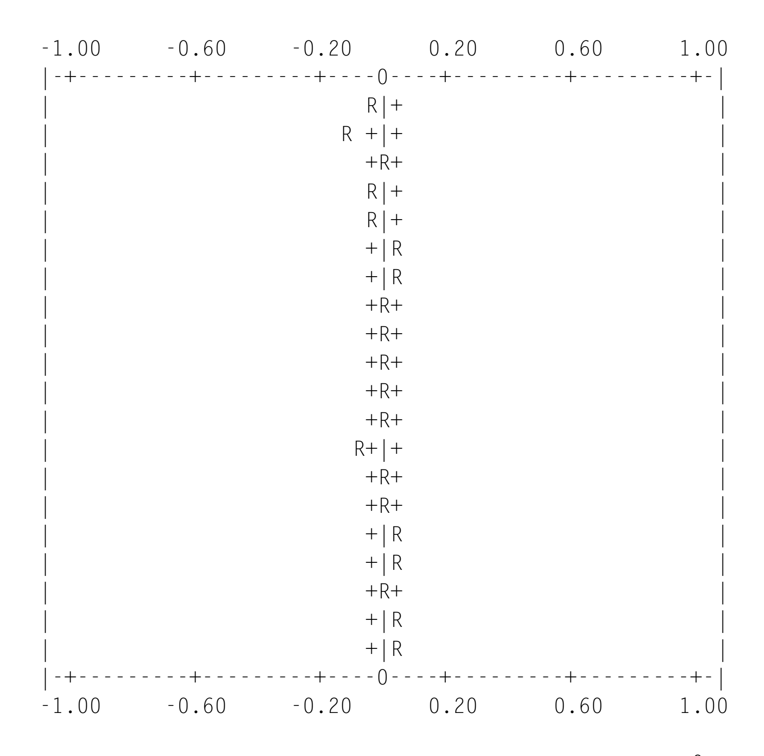
                     Autocorrelations

                                                     0    0 0

                                         Series (1-B) (1-B )  DOBL
                                                     Mean =   0.15738376E-03
                                           Standard Error =   0.57238708E-01
                | Lags|
Autocorrelations|      |-0.517E-01|-0.109    |-0.100E-01|-0.214E-01|-0.556E-01
Standard Errors | 1- 5| 0.375E-01| 0.376E-01| 0.380E-01| 0.380E-01| 0.381E-01
Q-statistics    |      | 1.91 10.4| 10.4 10.8| 13.0     |          |
Autocorrelations|      | 0.449E-01| 0.451E-01| 0.154E-01| 0.110E-01|-0.127E-01
Standard Errors | 6-10| 0.382E-01| 0.383E-01| 0.383E-01| 0.383E-01| 0.383E-01
Q-statistics    |      | 14.4     | 15.9     | 16.1     | 16.2     | 16.3
Autocorrelations|      | 0.495E-02|-0.127E-01|-0.710E-01|-0.116E-01| 0.142E-01
Standard Errors |11-15| 0.383E-01| 0.383E-01| 0.384E-01| 0.385E-01| 0.385E-01
Q-statistics    |      | 16.3     | 16.4     | 20.1     | 20.2     | 20.3
Autocorrelations|      | 0.261E-01| 0.227E-01|-0.457E-02| 0.371E-01| 0.248E-01
Standard Errors |16-20| 0.385E-01| 0.386E-01| 0.386E-01| 0.386E-01| 0.386E-01
Q-statistics    |      | 20.8     | 21.2     | 21.2     | 22.2     | 22.7
```

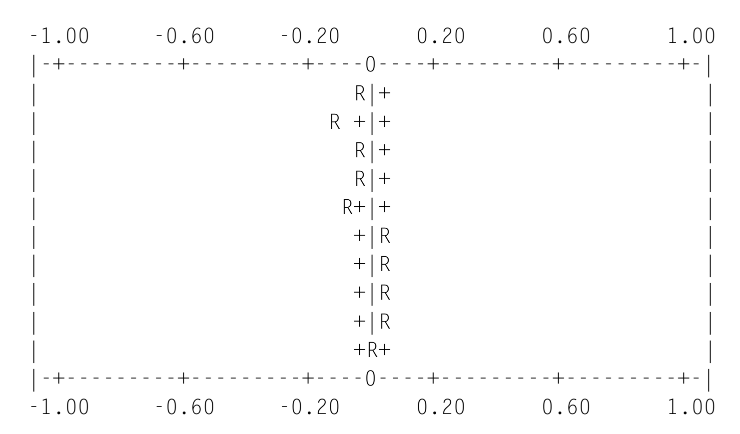



```
                       Partial Autocorrelations
                       ========================
                                           0     0 0
                                Series (1-B) (1-B )   DOBL
                   Standard Error of Autocorrelations =  0.37502930E-01
                 | Lags|
Partial Autocorrs| 1-5 |-0.517E-01|-0.112    |-0.224E-01|-0.360E-01|-0.637E-01
Partial Autocorrs| 6-10| 0.315E-01| 0.360E-01| 0.266E-01| 0.217E-01|-0.524E-02
Partial Autocorrs|11-15| 0.375E-01| 0.157E-03| 0.572E-01| 0.200E-02|-0.133
Partial Autocorrs|16-20|-0.120E-01|-0.200E-01|-0.260E-01|-0.450E-01|-0.390E-01
Partial Autocorrs|21-10|          |          |          |          |
```



```
                                            0    0 0
        Autocorrelation Function of the series (1-B) (1-B ) DOBL

-1.00     -0.60     -0.20      0.20      0.60      1.00
|-+---------+---------+----0----+---------+---------+-|
|                         R|+                         |
|                       R +|+                         |
|                         +R+                         |
|                         R|+                         |
|                         R|+                         |
|                         +|R                         |
|                         +|R                         |
|                         +R+                         |
|                         +R+                         |
|                         +R+                         |
|                         +R+                         |
|                         +R+                         |
|                        R+|+                         |
|                         +R+                         |
|                         +R+                         |
|                         +|R                         |
|                         +|R                         |
|                         +R+                         |
|                         +|R                         |
|                         +|R                         |
|-+---------+---------+----0----+---------+---------+-|
-1.00     -0.60     -0.20      0.20      0.60      1.00
```



```
                                            0     0 0
 Partial Autocorrelation Function of the series (1-B) (1-B ) DOBL

-1.00     -0.60     -0.20      0.20      0.60      1.00
|-+---------+---------+----0----+---------+---------+-|
|                         R|+                         |
|                       R +|+                         |
|                         R|+                         |
|                         R|+                         |
|                        R+|+                         |
|                         +|R                         |
|                         +|R                         |
|                         +|R                         |
|                         +|R                         |
|                         +R+                         |
|-+---------+---------+----0----+---------+---------+-|
-1.00     -0.60     -0.20      0.20      0.60      1.00
```

Sur les autocorrélogrammes simples et partiels, donc sur les graphiques des autocorrélations simples empiriques et des autocorrélations partielles empiriques, les *R* représentent les autocorrélations empiriques simples r_θ et partielles $\hat{\phi}_{\theta\theta}$. Sur chaque graphique, sont également dessinées les bornes d'un intervalle. Si un coefficient d'autocorrélation empirique (c'est-à-dire estimé) d'un certain ordre θ se trouve à l'intérieur de l'intervalle, on ne rejette pas l'hypothèse que le coefficient d'autocorrélation vrai correspondant est nul.

Ces graphiques sont précédés de tableaux donnant les valeurs calculées de ces coefficients d'autocorrélation empiriques, pour chaque retard (lag) allant de 1 à 20. Dans le tableau des autocorrélations simples empiriques r_θ (le premier tableau), on trouve, sous chaque coefficient d'autocorrélation estimé, une estimation de son écart type (standard error) et un test *Q*-statistic, qui est celui de Ljung et Box [LJU 1978].

Le graphique des autocorrélations simples empiriques r_θ (les valeurs estimées des ρ_θ) suggère que seuls les coefficients d'autocorrélation ρ_1 et ρ_2, voire ρ_2 uniquement, sont non nuls. Le modèle ARMA adéquat serait donc un *MA*(2). Le graphique des autocorrélations partielles $\hat{\phi}_{\theta\theta}$ (les valeurs estimées des $\phi_{\theta\theta}$) suggère une sinusoïde s'amenuisant, ce qui est possible pour un *MA*(2). On peut donc estimer un modèle *MA*(2) pour DOBL, grâce au programme suivant :

```
freq n;
smpl 1 712;
read(file='inr.xls');
smpl 2 712;
dobl=obl-obl(-1);
bjest(nma=2,nar=0,constant,nocumplo,noplot) dobl;
```

Après la lecture des données de la série OBL dans le fichier INR.XLS, et la modification de l'échantillon pour le calcul de la première différence DOBL, l'instruction `BJEST` lance l'estimation d'un modèle ARMA pour DOBL. Plus précisément, les options `nma=2,nar=0` demandent d'estimer un *MA*(2). En effet, l'option `nma=2` exige deux coefficients autorégressifs, tandis que l'option `nar=0` ne nécessite aucun coefficient de moyenne mobile. `constant` donne à ce modèle une constante μ, et les autres options empêchent l'insertion d'une série de graphiques dans les résultats. La sortie est la suivante :

```
Autocorr     0.044909      0.044762         PROGRAM
LINE ****************************************************************
|      1  freq n;
|      2  smpl 1 712;
|      3  read(file='inr.xls');
|      4  smpl 2 712;
|      5  dobl=obl-obl(-1);
|      6  bjest(nma=2,nar=0,constant,nocumplo,noplot) dobl;
         EXECUTION

Current sample:  1 to 712
Current sample:  2 to 712

Box-Jenkins procedures
Procedure   BJEST

Working space used: 9217
```

```
                                STARTING VALUES

                 THETA1       THETA2       CONSTANT
VALUE        0.051695       0.10877     0.00015738
F=  -1031.4    FNEW=  -1031.5    ISQZ=  0 STEP=  1.0000    CRIT= 0.18165
F=  -1031.5    FNEW=  -1031.5    ISQZ=  0 STEP=  1.0000    CRIT= 0.84606E-03
F=  -1031.5    FNEW=  -1031.5    ISQZ=  0 STEP=  1.0000    CRIT= 0.53143E-05

CONVERGENCE ACHIEVED AFTER 3 ITERATIONS

6 FUNCTION EVALUATIONS.

                         Results of Box-Jenkins Estimation

Dependent variable: DOBL
Current sample:  2 to 712
Number of observations:  711

 Mean of dependent variable = .157384E-03
Std. dev. of dependent var. = .057239
   Sum of squared residuals = 2.28724
      Variance of residuals = .323057E-02
   Std. error of regression = .056838
                  R-squared = .016754
         Adjusted R-squared = .013976
    Durbin-Watson statistic = 1.99433
  F-statistic (zero slopes) = 6.02163
 Log of likelihood function = 1031.46

                         Standard
Parameter  Estimate        Error       t-statistic
THETA1     .064547       .037309       1.73004
THETA2     .119353       .037310       3.19891
CONSTANT   .194120E-03   .174057E-02   .111527

Standard Errors computed from quadratic form of analytic first
derivatives (Gauss)
                 Autocorrelations of the Residuals
                1             2            3            4             5
Autocorr    0.0028068     0.0054788    -0.016550    -0.019274     -0.050715
Q-stat      0.0056250      0.027088      0.22321      0.48959       2.33643
P-value       .             .            .          0.48411       0.31092

                6             7            8            9            10
Autocorr    0.044909       0.044762     0.022727     0.016507     -0.011537
Q-stat       3.78664        5.22945      5.60193      5.79871       5.89497
P-value      0.28544        0.26455      0.34690      0.44611       0.55207

               11            12           13           14            15
Autocorr  -0.0033513     -0.019970    -0.071881    -0.014134      0.010081
Q-stat       5.90310        6.19233      9.94497     10.09026      10.16428
P-value      0.65809        0.72053      0.44533      0.52229       0.60155

               16            17           18           19            20
Autocorr    0.027060       0.030335    0.0057656     0.040393      0.027058
```

```
Q-stat      10.69839      11.37057    11.39489      12.59014      13.12728
P-value      0.63607       0.65671     0.72413       0.70246       0.72762
```

Les résultats semblent indiquer qu'un modèle $MA(2)$ est une bonne description du vrai processus stochastique dont DOBL est une réalisation observée :

$$\begin{aligned} DOBL_t &= \text{CONSTANT} + v_t - \text{THETA1}v_{t-1} - \text{THETA2}v_{t-2} \\ &= .194120\text{E} - 03 + v_t - .064547v_{t-1} - .119353v_{t-2} \end{aligned}$$

où v_t est un résidu calculé, CONSTANT la valeur estimée de μ, et THETA1 et THETA2 les valeurs estimées de θ_1 et θ_2.

Les t-stats permettent de tester l'hypothèse qu'un coefficient vrai vaut 0, en étant distribuées approximativement par une $N(0, 1)$ si c'est vrai. À 5 %, la valeur critique est de 1,96 et l'on ne peut donc rejeter l'hypothèse que $\theta_1 = 0$. Cela ne veut pas dire qu'on est sûr que $\theta_1 = 0$. Cela signifie que c'est une possibilité que l'on ne peut rejeter, parmi une infinité d'autres valeurs qui ne seraient pas rejetées non plus. Ces valeurs de θ_1 non rejetées appartiennent à l'intervalle de confiance allant de « THETA1 $-$ 1,96$\times$ écart type estimé de THETA1 » à « THETA1 $+$ 1,96$\times$ écart type estimé de THETA1 », qui a 90 chances sur 100 de contenir la valeur « vraie inconnue » de θ_1. Par conséquent :

$$P\big(064547 - 1{,}96(0.037309) < \theta_1 < 064547 + 1{,}96(0.037309)\big) = 0.9$$

Le terme d'erreur vrai semble être un bruit blanc d'après les tests Q-stat (de Ljung et Box [LJU 1978]) sur les autocorrélations des résidus. Chaque test Q-stat vérifie en effet l'hypothèse que les autocorrélations des ordres allant de 1 à une certaine valeur i (indiquée au-dessus du test dans le tableau) sont nulles, donc que le résidu n'est pas corrélé avec ses i valeurs précédentes. Si l'hypothèse est vraie, le test est distribué selon une Chi-2 à I degrés de liberté. Les probabilités critiques indiquées étant supérieures à 0,05 pour toutes les valeurs de i présentées, on ne rejette pas l'hypothèse d'absence d'autocorrélation du résidu v, au seuil de 5 %.

Si l'on veut utiliser ce modèle pour effectuer des extrapolations de la série sur des périodes ultérieures, on exécute la commande TSP spéciale `BJFRCST`. Par exemple, si un analyste souhaite, le 22/9/2000, prévoir les taux d'intérêt des 10 prochains jours ouvrables, et dispose des observations historiques des 712 derniers jours, il envoie à TSP le programme suivant :

```
freq n;
smpl 1 712;
read(file='inr.xls');
smpl 2 712;
dobl=obl-obl(-1);
bjest(nma=2,nar=0,constant,nocumplo,noplot) dobl;
bjfrcst(constant,nar=0,nma=2,retrieve,nhoriz=10,orgbeg=712,print) dobl;
smpl 712 712;
oblf=obl;
smpl 713 722;
oblf=@fit+oblf(-1);
print oblf;
```

Les options `constant,nar=0,nma=2` demandent que la prévision soit utilisée avec un modèle $MA(2)$ avec constante. `retrieve` indique que le logiciel n'a pas besoin d'estimer ce

modèle mais qu'il peut tout simplement récupérer les coefficients estimés par la commande `BJEST` précédente. `nhoriz=10` exige de prévoir les valeurs pour 10 périodes successives. `orgbeg=712` indique la date d'origine à partir de laquelle il faut calculer les extrapolations.

Le logiciel calcule des prévisions sur DOBL (la variation des taux d'intérêt) et les stocke en mémoire sous le nom `@fit`. Les instructions suivantes transforment ces prévisions sur DOBL en prévisions (appelées `oblf`) sur le niveau OBL des taux d'intérêt.

On retrouve alors le même fichier de résultats TSP que précédemment, avec en plus les résultats suivants de l'instruction `BJFRCST` ajoutée au programme :

```
Box-Jenkins procedures
Procedure   BJFRCST

                        OPTIONS FOR THIS ROUTINE

CONBOUND = 0.95000      CONSTANT = TRUE          NAR      = 0
NBACK    = 100          NDIFF    = 0             NHORIZ   = 10
NLAG     = 20           NMA      = 2             NSAR     = 0
NSDIFF   = 0            NSMA     = 0             NSPAN    = 0
ORGBEG   = 712          ORGEND   = 712           PLOT     = FALSE
PRINT    = TRUE         RETRIEVE = TRUE

TIME SERIES :DOBL
STANDARD ERROR OF THE DISTURBANCE = 0.056838
THETA(B)
                                2
      1 - 0.064547 B - 0.11935 B

CONSTANT =0.00019412
THETA*(B)
                                2
      1 - 0.064547 B - 0.11935 B

                    FORECAST STANDARD ERRORS AND PSI WEIGHTS
                STD ERR          PSI
       1       0.056838    -0.064547
       2       0.056956     -0.11935
       3       0.057359      0.00000
       4       0.057359      0.00000
       5       0.057359      0.00000
       6       0.057359      0.00000
       7       0.057359      0.00000
       8       0.057359      0.00000
       9       0.057359      0.00000
       10      0.057359      0.00000

               Forecasts and 95\% Confidence Bounds (Origin = 712)

              Lowr Bnd      Forecast       Uppr Bnd        Actual        Error
 712        -0.0012999   -0.0012999     -0.0012999     -0.0012999      0.00000
 713          -0.10731    0.0040861        0.11549           .            .
 714          -0.11011    0.0015212        0.11315           .            .
 715          -0.11223    0.00019412       0.11262           .            .
 716          -0.11223    0.00019412       0.11262           .            .
 717          -0.11223    0.00019412       0.11262           .            .
```

```
 718          -0.11223    0.00019412      0.11262        .            .
 719          -0.11223    0.00019412      0.11262        .            .
 720          -0.11223    0.00019412      0.11262        .            .
 721          -0.11223    0.00019412      0.11262        .            .
 722          -0.11223    0.00019412      0.11262        .            .


Current sample:  712 to 712
Current sample:  713 to 722
NOTE: Dynamic GENR for OBLF

                 OBLF
713          5.44999
714          5.45151
715          5.45170
716          5.45190
717          5.45209
718          5.45228
719          5.45248
720          5.45267
721          5.45287
722          5.45306
```

Le logiciel calcule des extrapolations des valeurs de DOBL (colonne Forecast) pour toutes les périodes demandées, et indique pour chaque période les bornes inférieure (colonne Lowr Bnd) et supérieure (colonne Uppr Bnd) d'un intervalle qui a une probabilité de 95 % de contenir la valeur « vraie encore inconnue » de DOBL à la période considérée. Ensuite, les extrapolations de DOBL sont transformées en extrapolations sur OBL (colonne OBLF).



Références bibliographiques

[AKA 1974] H. Akaike, A New Look at the Statistical Identification Model, dans *IEEE : Trans. Auto. Control*, 19, p. 716–723, 1974.

[AKA 1973] H. Akaike, Information Theory and the Extension of the Maximum Likelihood Principle, dans *Proceeding of the Second International Symposium on Information Theory*, eds B.N. Petrov and F. Csaki, Budapest, p. 267–281, 1973.

[BIE 1997] H.J. Bierens, Testing the unit root hypothesis against nonlinear trend stationarity, with an application to the price level and interest rate in the U.S., dans *Journal of Econometrics*, 81, p. 29–64, 1997.

[BIE 1996] H.J. Bierens, *Topics in Advanced Econometrics*, Cambridge University Press (paper back edition), 1996.

[BIE 1993a] H.J. Bierens, Higher Order Autocorelations and the Unit Root Hypothesis, dans *Journal of Econometrics*, 57, p. 137–160, 1993.

[BIE 1993b] H.J. Bierens, S. Guo, Testing Stationarity and Trend Stationarity Against the Unit Root Hypothesis, dans *Econometric Reviews*, 12, p. 1–32, 1993.

[BOX 1976] G.P. Box, G.M. Jenkins, *Times series forecasting and control*, Holden-Day, New York, 1976.

[CAM 1997] Campbell, Lo, MacKinlay, *The Econometrics of Financial Markets*, Princeton University, 1997.

Davidson, R. and J.G. McKinnon, 1993, Estimation and Inference in Econometrics, Oxford University Press, Oxford.

[DIC 1979] D.A. Dickey, W.A. Fuller, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, dans *Journal of the American Statistical Association*, 74, p. 427–431, 1979.

[FUL 1996] W.A. Fuller, *Introduction to Statistical Time Series*, 2nd ed., New York, John Wiley, 1996.

[HAR 1981] A.C. Harvey, *The econometric analysis of time series*, Philip Allan, London, 1981.

[KEN 1983] M. Kendall, A. Stuart, J.K. Ord, *The advanced theory of statistics*, vol 3, Charles Griffin, London, 1983.

[KWI 1992] D. Kwiatkowski, P. Phillips, P. Schmidt, Y. Shin, Testing the Null of Stationarity Against the Alternative of a Unit Root, dans *Journal of Econometrics*, 54, p. 159–178, 1992.

[LJU 1978] G.M. Ljung, G.P. Box, On a measure of lack of fit in time series models, dans *Biometrika*, 66, p. 297–303, 1978.

[MAC 1993] J.G. MacKinnon, Approximate Asymptotic Distribution Function for Unit Root and Cointegration Tests, dans *Journal of Business and Economic Statistics*, 1994.

[NEW 1987] W.K. Newey, K.D. West, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, dans *Econometrica*, 55, p. 703–708, 1987.

[PAN 1994] S.G. Pantula, G. Gonzales Farias, W.A. Fuller, A Comparison of Unit Root Test Criteria, dans *Journal of Business and Economic Statistics*, p. 449, 1994.

[PER 1989] P. Perron, The Great Crash, the Oil Shock and the Unit Root Hypothesis, dans *Econometrica*, p. 1361–1401, 1989.

[PES 1997] H. Pesaran, B. Pesaran, *Microfit 4.0 Windows Version*, Oxford University Press, Oxford, 1997.

[PHL 1987] Phlips, Blomme, Vanden Berghe, Dor, *Analyse chronologique*, éditions De Boek, 1987.

[PHI 1988] P.C.B. Phillips, P. Perron, Testing for a Unit Root in Time Series Regression, dans *Biometrica*, 75, p. 335–346, 1988.

[PHI 1987] P.C.B. Phillips, Time Series Regression with a Unit Root, dans *Econometrica*, 55, p. 277–301, 1987.

[SAI 1991] S.E. Said, Unit Root Test for Time Series Data with a Linear Time Trend, dans *Journal of Econometrics*, 47, p. 285–303, 1991.

[SAI 1984] S.E. Said, D.A. Dickey, Testing for Unit Roots in Autoregresive Moving Average of Unknown Order, dans *Biometrika*, 71, p. 599–607, 1984.

[SCH 1978] G. Schwarz, Estimating the Dimension of a Model, dans *Annals of Statistics*, 6, p. 461–464, 1978.

[SCH 1989] G.W. Schwert, Tests for Unit Roots : a Monte Carlo Investigation, dans *Journal of Business and Economic Statistics*, p. 147, 1989.

[SHI 1998] Shin, Dong Wan, Fuller, Wayne, Unit root tests based on unconditional maximum likelihood estimation for the autoregressive moving average, dans *Journal of Time Series Analysis*, 19 #5, p. 591–599, 1998.

[STO 1994] J.M. Stock, *Unit Roots, Structural Breaks and Trends, in Handbook of Econometrics*, eds R.F. Engle and D. McFadden, North Holland, p. 2738–2841, 1994.

Chapitre 6

Variables intégrées, modèles *VAR* et cointégration

Variables intégrées, modèles *VAR* et cointégration

Ce chapitre définit d'abord la cointégration entre variables intégrées du même ordre. Il décrit également les tests de cointégration d'Engle et Granger, avant d'aborder les estimateurs modifiés des coefficients d'une relation linéaire entre variables intégrées. Il détaille ensuite les propriétés de cointégration dans une modélisation *VAR* et les tests de Johansen. Grâce à cette structure d'analyse, on peut approfondir la modélisation en définissant les notions de *VAR* conditionnel à correction d'erreur et de modèle structurel à correction d'erreur.

1 Problèmes économétriques liés aux variables intégrées

Plusieurs problèmes se posent lorsqu'on veut vérifier l'existence d'une relation linéaire entre plusieurs variables dont certaines ont une tendance stochastique (donc une racine unitaire), et lorsqu'on veut estimer les paramètres de cette relation supposée :

- Même si, dans la réalité, aucune relation linéaire ne lie ces variables, une estimation par MCO peut donner des résultats qui font croire faussement qu'une telle relation existe et qu'elle est importante (R^2 élevé, t-stats significatifs...). C'est le phénomène bien connu de *régression fallacieuse* ou *régression factice* (*spurious regression* [1] en anglais) [GRA 1974] et [PHI 1986]. En fait, l'existence d'une réelle relation à long terme entre des variables intégrées est soumise à certaines conditions, appelées *cointégration entre les variables intégrées*. En d'autres termes, si les variables sont intégrées (ce que l'on vérifie avec les tests de racine unitaire), il faut vérifier leur éventuelle cointégration pour savoir si elles entretiennent réellement une relation à long terme. Ces conditions de cointégration sont détaillées dans ce chapitre, ainsi que des méthodes permettant de vérifier empiriquement une éventuelle cointégration entre des variables intégrées observées.
- Même en cas de cointégration entre la variable dépendante et les variables explicatives, les estimateurs des moindres carrés ordinaires convergent « trop vite » pour être distribués d'après une loi normale. Par conséquent, les tests d'hypothèse usuels ne suivent pas les lois Student ou asymptotiquement normales : l'inférence statistique classique (voir chapitres 2, 3 et 4) ne s'applique pas. D'autres techniques d'estimation permettent toutefois de générer des tests d'hypothèse sur les coefficients cointégrants, qui utilisent des distributions connues.

2 Définition de la cointégration

Des processus stochastiques $X_1, X_2 \ldots X_w$ **intégrés du même ordre d** sont cointégrés s'il existe une combinaison linéaire de ces processus qui est intégrée d'un ordre inférieur à d.

Il faut donc qu'il existe une valeur $b > 0$ et des valeurs $\beta_1, \beta_2 \ldots \beta_w$ vérifiant :

- $\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_w X_w$ est $I(d-b)$.
- Chaque variable $X_1, X_2 \ldots X_w$ est $I(d)$.

Remarque

En pratique, le cas le plus fréquent est celui où toutes les variables sont intégrées d'ordre 1 ($d = 1$), et où il existe une combinaison linéaire de celles-ci, stationnaire et donc intégrée d'ordre 0 ($b = 1$ et $d - b = 1 - 1 = 0$).

La définition générale permet toutefois de tenir compte d'autres cas possibles, quoique rares. Par exemple, des variables peuvent être chacune $I(2)$ et cointégrées en raison de l'existence d'une combinaison linéaire de celles-ci de type $I(1)$ (dans ce cas, $d = 2$, $b = 1$ et $d - b = 1$) ou $I(0)$ (dans ce cas, $d = 2$, $b = 2$ et $d - b = 0$).

$\beta = (\beta_1, \beta_2 \ldots \beta_w)'$ est le *vecteur cointégrant* (ou de cointégration). $\beta' X_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_w X_{wt}$ est l'*expression cointégrée*.

Il peut exister plusieurs vecteurs cointégrants linéairement indépendants pour les mêmes w variables. Le nombre de ces vecteurs est noté r. Dans tous les cas, $r < w$; en d'autres termes on ne peut avoir plus de $w - 1$ vecteurs de cointégration linéairement indépendants.

1. On doit la découverte de ce phénomène à Yule [YUL 1926].



Toute combinaison linéaire de vecteurs de cointégration pour des variables est aussi un vecteur de cointégration pour ces variables. Tout multiple d'un vecteur de cointégration est également un vecteur de cointégration. Bien entendu, ce nouveau vecteur est linéairement dépendant du premier.

Une même direction vectorielle est représentée par une infinité de vecteurs, tous linéairement dépendants, tous combinaisons linéaires les uns des autres. L'espace de vecteurs de cointégration est donc infini, même si $r = 1$, c'est-à-dire même s'il existe un seul vecteur de cointégration linéairement indépendant. Par « r vecteurs de cointégration linéairement indépendants », on entend donc « r directions vectorielles de cointégration différentes », chacune d'elles étant représentée par une infinité de vecteurs qui sont des multiples les uns des autres. Pour représenter de manière unique chacune de ces r directions vectorielles de cointégration, il faut appliquer une règle de normalisation arbitraire : par exemple, on peut les représenter toutes par leur vecteur dont le premier coefficient (le coefficient de la première variable) vaut 1, ou encore par leur vecteur dont le deuxième coefficient (celui de la deuxième variable) vaut 1,... On peut aussi représenter la première direction vectorielle (respectivement, la deuxième, la troisième...) par son vecteur dont le premier (respectivement, le deuxième, le troisième...) coefficient vaut 1(et ainsi de suite).

Quand $r > 1$, même si chacune des r directions vectorielles a été normalisée et est donc représentée par un seul vecteur normalisé, toute combinaison linéaire de ces r vecteurs est aussi un vecteur de cointégration. Avoir normalisé les directions initiales ne suffit donc pas à limiter le nombre de vecteurs à r. Pour sélectionner seulement r vecteurs, il faut en plus appliquer une règle arbitraire d'identification, généralement fournie par des restrictions sur les coefficients suggérées par la théorie économique, financière, marketing...

L'exemple le plus fréquent est celui de w variables $X_1, X_2 \ldots X_w$, toutes $I(1)$ et cointégrées s'il existe une combinaison linéaire $\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_w X_w$ stationnaire, donc une expression $\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_w X_w I(0)$. En pareil cas, chaque variable individuelle est non stationnaire avec une tendance stochastique puisqu'elle est $I(1)$, et a donc tendance à s'éloigner davantage de ses propres conditions initiales, avec de larges fluctuations. Toutefois, étant cointégrées, ces variables divergent ensemble : elles ont tendance à ne pas trop s'éloigner les unes des autres, puisqu'une fonction de ces variables reste stationnaire. Une force économique les relie donc à long terme, de manière qu'elles ne s'écartent pas durablement d'une relation d'équilibre.

La combinaison linéaire $\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_w X_w$ représente la déviation par rapport à une relation d'équilibre entre les variables. La cointégration implique que cette déviation est stationnaire : elle fluctue autour d'une espérance constante, avec une dispersion constante. Cette déviation n'a donc pas tendance à diverger, à devenir trop grande : les variables ne s'écartent pas durablement de leur relation d'équilibre.

Exemple

À long terme, la consommation agrégée réelle doit être proportionnelle au revenu disponible réel des ménages, d'après la théorie économique. À l'équilibre, il faut donc que :

$$C = \mathrm{A}Y$$

où C est la consommation totale réelle, A une constante et Y le revenu disponible total réel. En effet, la théorie économique montre que, logiquement, le taux de consommation doit être constant à l'équilibre :

$$\frac{C}{Y} = \mathrm{A}$$

ou encore que le taux d'épargne doit l'être aussi :

$$\frac{Y - C}{Y} = 1 - \mathrm{A}$$

Par conséquent, la relation suivante est vérifiée à l'équilibre :

$$\ln(C) = \mathrm{A}' + \ln(Y)$$

où $\mathrm{A}' = \ln(\mathrm{A})$. Bien entendu, cette relation prévaut sur le long terme, mais pas à court terme, ni à chaque période t. À court terme, la consommation logarithmique $\ln(C_t)$ peut s'écarter de son équilibre $\mathrm{A}' + \ln(Y_t)$ d'un certain montant u_t :

$$\ln(C_t) = \mathrm{A}' + \ln(Y_t) + u_t \quad \text{pour tout } t$$

Mais u_t ne peut devenir indéfiniment grand : il doit fluctuer autour de 0, de manière stationnaire, pour permettre à la relation d'équilibre de prévaloir approximativement sur le long terme. Les tests de racine unitaire indiquent que les variables $\ln(C_t)$ et $\ln(Y_t)$ liées par la relation précédente sont intégrées. Cette équation s'écrit encore :

$$\ln(C_t) - \ln(Y_t) = \mathrm{A}' + u_t \quad \text{pour tout } t$$

où A' est une constante et u_t un processus stochastique stationnaire. La théorie économique implique ainsi que la combinaison linéaire $\ln(C_t) - \ln(Y_t)$ est stationnaire, donc que $\ln(C_t)$ et $\ln(Y_t)$ sont cointégrées, avec $\beta_1 = 1$ et $\beta_2 = -1$. Les tests de cointégration qui suivent permettent de vérifier concrètement, sur les données de la réalité, si cette implication de la théorie économique est vraie.

3 Test de cointégration d'Engle et Granger

3.1 Principe

Le test d'Engle et Granger est une méthode de vérification de l'existence d'une relation de cointégration entre des variables intégrées et d'estimation de cette relation. Cette méthode est valable sous l'hypothèse arbitraire qu'il existe un seul vecteur de cointégration entre les variables utilisées, et que $b = d$.

Le raisonnement est le suivant : en cas de cointégration entre $X_1, X_2 \ldots X_w$, il existe des valeurs $\beta_1, \beta_2 \ldots \beta_w$ telles que :

$$\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_w X_w \quad \text{est } I(0) \tag{6.1}$$

et donc telles que :

$$X_1 + \frac{\beta_2}{\beta_1} X_2 + \cdots + \frac{\beta_w}{\beta_1} X_w \quad \text{est aussi } I(0) \tag{6.2}$$

et donc telles que :

$$X_1 + \tilde{\beta}_2 X_2 + \cdots + \tilde{\beta}_w X_w \quad \text{est aussi I}(0) \tag{6.3}$$

où les coefficients sont normalisés : $\tilde{\beta}_1 = \frac{\beta_1}{\beta_1} = 1, \tilde{\beta}_2 = \frac{\beta_2}{\beta_1} \ldots \tilde{\beta}_w = \frac{\beta_w}{\beta_1}$.

Tout processus $I(0)$ est forcément égal à une constante (son espérance) + un processus $I(0)$ d'espérance nulle. La stationnarité de $X_1 + \tilde{\beta}_2 X_2 + \cdots + \tilde{\beta}_w X_w$ implique que :

$$X_{1t} + \tilde{\beta}_2 X_{2t} + \cdots + \tilde{\beta}_w X_{wt} = \mu + u_t \tag{6.4}$$



où μ est une constante (non aléatoire) et u un processus stochastique stationnaire $I(0)$ d'espérance nulle. u n'est pas nécessairement un bruit blanc. Il peut être autocorrélé, dès lors qu'il est stationnaire. La cointégration implique donc que :

$$X_{1t} = \mu + (-\tilde{\beta}_2)X_{2t} + \cdots + (-\tilde{\beta}_w)X_{wt} + u_t \quad (6.5)$$

où les coefficients cointégrants normalisés $\tilde{\beta}_2 \ldots \tilde{\beta}_w$ sont les seuls pour lesquels u possède les propriétés équivalentes suivantes :

- Il est stationnaire $I(0)$.
- Il n'a pas de tendance stochastique.
- Il n'a pas de racine unitaire.
- Il a une variance constante (voir chapitre 5).

Les coefficients cointégrants de l'équation précédente sont susceptibles d'être estimés par MCO. Cette méthode cherche les valeurs des coefficients qui minimisent la somme des carrés des résidus. Minimiser cette somme revient à chercher les valeurs des coefficients qui minimisent la variance estimée des u_t puisque $\hat{\sigma}_u^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-w}$ (voir chapitre 4). Or la variance de u est finie uniquement pour les coefficients cointégrants $\tilde{\beta}_2 \ldots \tilde{\beta}_w$. Pour d'autres valeurs (d'autres combinaisons linéaires des X_{it}), u est $I(1)$ et a une variance infinie. Par conséquent, en cas de cointégration, et donc s'il existe des coefficients cointégrants $-\tilde{\beta}_2 \ldots - \tilde{\beta}_w$ pour lesquels le terme d'erreur u de l'équation précédente a une variance constante, l'estimation de cette équation par MCO doit nécessairement fournir des valeurs estimées de ces coefficients cointégrants.

Tester la cointégration entre les variables revient donc à estimer par MCO une équation linéaire où l'une de ces variables est régressée sur les autres et à tester si le résidu a une racine unitaire. Si l'on ne rejette pas l'hypothèse de racine unitaire pour le résidu estimé, les variables de l'équation ne sont pas cointégrées ; si on la rejette, elles le sont.

Toutefois, les variables étant intégrées, l'inférence statistique classique ne s'applique pas. Les t-ratios des coefficients estimés n'ont pas une distribution Student pour des petits échantillons, ni une distribution gaussienne pour de grands échantillons, sous l'hypothèse de nullité du coefficient vrai correspondant. Les probabilités critiques des t-ratios, indiquées par défaut par les logiciels et calculées à partir d'une distribution Student, ne sont donc pas fiables. À ce niveau, il est donc impossible de réaliser des tests d'hypothèse et des intervalles de confiance, pour vérifier s'ils admettent des valeurs des coefficients plausibles économiquement. Il est donc nécessaire de estimer de nouveau la relation proposée au moyen des estimateurs Fully Modified (FM) de Phillips et Hansen [PHI 1990a], qui fournissent des valeurs estimées différentes, accompagnées d'écarts types permettant une inférence statistique classique (voir section 4.2).

3.2 Cointégration et mécanisme à correction d'erreur

Selon le théorème de représentation de Granger, en cas de cointégration entre des variables intégrées d'ordre 1, l'évolution de chacune d'elles est régie nécessairement par un *modèle à correction d'erreur*. Pour chaque variable cointégrée, un modèle linéaire existe donc, où la variation de la variable est une fonction de ses variations passées et des variations

passées des autres variables, ainsi que de la valeur passée de l'expression cointégrée :

$$X_1, X_2 \ldots X_w \text{ cointégrées} \Leftrightarrow \forall i = 1, \ldots, w \; :$$

$$\Delta X_{it} = \mu_i + \sum_{j=1}^{w} \sum_{\theta=1}^{r} \varphi_{ij\theta} \Delta X_{jt-\theta} + \lambda_i \left(X_{1t-1} - \tilde{\beta}_2 X_{2t-1} - \cdots - \tilde{\beta}_w X_{wt-1} \right) + u_{it} \quad (6.6)$$

où u_{it} est un processus stochastique de type bruit blanc. Un tel modèle est dit « à correction d'erreur » (on parle aussi de mécanisme à correction d'erreur).

L'expression cointégrée retardée, notée $\left(X_{1t-1} - \tilde{\beta}_2 X_{2t-1} - \cdots - \tilde{\beta}_w X_{wt-1} \right)$, représente l'écart (ou l'« erreur ») par rapport à la relation d'équilibre[1] à la période $t - 1$. Le paramètre λ_i correspond donc au taux de réaction de la variable X_i à l'écart précédent par rapport à l'équilibre. λ_i mesure ainsi l'intensité avec laquelle la variable X_i varie en t pour corriger l'erreur de la période $t - 1$ par rapport à l'équilibre. C'est pourquoi on parle de modèle à correction d'erreur.

4 Estimateurs modifiés

4.1 Intérêt pratique

En cas de cointégration entre des variables, l'estimation par MCO d'une équation linéaire, expliquant une de ces variables par les autres, fournit un estimateur des coefficients β dont la distribution asymptotique n'est pas une loi normale. Il en résulte que la distribution asymptotique d'un test de Wald de restrictions sur β n'a pas une distribution classique mais dépend, sous l'hypothèse nulle, de plusieurs paramètres de nuisance, et plus précisément :

- D'un biais qui est fonction des liens entre le terme d'erreur et les variables contemporaines et retardées.
- Des variances du terme d'erreur et d'un mouvement brownien U particulier, associé à l'accumulation des termes d'erreur successifs, qui ne sont égales que si le terme d'erreur n'est pas autocorrélé.
- Des covariances entre deux mouvements browniens U et B, associés à l'accumulation des termes d'erreur successifs et à l'accumulation des variations successives des variables explicatives. Ces covariances ne sont nulles que si les variables explicatives sont strictement exogènes pour l'équation.

L'intérêt pratique est évidemment de parvenir à vérifier des restrictions au moyen de tests ayant une distribution asymptotique classique. Il faut donc corriger le test de Wald pour que la version corrigée ait une distribution asymptotique classique. Les deux premiers problèmes cités précédemment peuvent être résolus par l'extension des méthodes de Park et Phillips (1988). Le troisième peut être réglé à l'aide d'une méthode proposée par Phillips et Hansen [PHI 1990a], qui permet d'obtenir un nouvel estimateur de β, différent de l'estimateur des MCO. Ce nouvel estimateur a une distribution asymptotique normale conditionnellement à B. Il en résulte qu'une fonction quadratique de cet estimateur a une distribution Chi-2. Par conséquent, les tests standard de Wald, du rapport de vraisemblance et du multiplicateur de Lagrange, obtenus à partir du nouvel estimateur (dit « modifié »), ont également une distribution asymptotique Chi-2.

---

1. Le mot « erreur » n'a pas le même sens que dans l'expression « terme d'erreur », où il est synonyme de « perturbation ».

4.2 Formules des estimateurs

La méthode de Phillips et Hansen est appropriée pour estimer et tester les paramètres d'une équation linéaire reliant des variables $I(1)$ cointégrées $Y, X_2 \dots X_k$:

$$Y_t = \mu + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \tag{6.7}$$

où chaque variable explicative est telle que :

$$\Delta X_{it} = \mu_i + v_{it} \quad \forall t = 2, \dots, n \quad \forall i = 2, \dots, k \tag{6.8}$$

et où le vecteur formé par le terme d'erreur u_t de l'équation linéaire et les différents v_{it} est stationnaire et a une matrice de variance et de covariance finie et définie positive.

Le calcul de l'estimateur FM des β_i par Phillips et Hansen se fait en deux étapes :

1. On corrige y_t pour l'interdépendance à long terme entre u_t et les v_{it}. On estime d'abord l'équation de Y_t par MCO et l'on forme le vecteur $\hat{\xi}_t = \begin{pmatrix} \hat{u}_t \\ \hat{\nu}_{2t} \\ \vdots \\ \hat{\nu}_{kt} \end{pmatrix}$ où $\hat{\nu}_{it} = \Delta X_{it} - \hat{\mu}_i$

   et $\hat{\mu}_i = \dfrac{\sum_{t=2}^{n} \Delta X_{it}}{n-1} \quad \forall i = 2, \dots, k.$

   Un estimateur convergent de la matrice de variance et de covariance de $\hat{\xi}_t$ est fourni par :

$$\hat{\Omega} = \hat{\Sigma} + \hat{\Lambda} + \hat{\Lambda}' = \begin{pmatrix} \hat{\Omega}_{11} & \hat{\Omega}_{12} \\ \hat{\Omega}_{21} & \hat{\Omega}_{22} \end{pmatrix} \tag{6.9}$$

   où $\hat{\Sigma} = \dfrac{\sum_{t=2}^{n} \hat{\xi}_t \hat{\xi}_t'}{n-1}$, $\hat{\Lambda} = \sum_{j=1}^{m} \omega(j, m) \hat{\Gamma}_j$, $\hat{\Gamma}_j = \dfrac{\sum_{t=1}^{n-j} \hat{\xi}_t \hat{\xi}_{t+j}'}{n}$ et où $\omega(j, m)$ est une fenêtre de retards avec une troncature m. Il existe plusieurs formes de fenêtre :

   – la fenêtre uniforme :

$$w(j, m) = 1 \quad \text{pour tout } j = 0 \dots m$$

   – la fenêtre de Bartlett :

$$w(j, m) = 1 - \frac{j}{m} \quad \text{pour tout } j = 0 \dots m$$

   – la fenêtre de Parzen :

$$w(j, m) = 1 - 6\left(\frac{j}{m}\right)^2 + 6\left(\frac{j}{m}\right)^3 \quad \text{si } 0 \leqslant j \leqslant \frac{m}{2}$$

   et

$$w(j, m) = 2\left(1 - \frac{j}{m}\right)^2 \quad \text{si } m \geqslant j > \frac{m}{2}$$

   – la fenêtre de Tukey :

$$w(j, m) = \frac{1}{2}\left\{1 + \cos\left(\frac{\pi j}{m}\right)\right\} \quad \text{pour tout } j = 0 \dots \frac{m}{2}$$

   On choisit la valeur de m arbitrairement. $\hat{\Omega}_{11}$ est de type 1×1, $\hat{\Omega}_{12}$ de type $1 \times k$, $\hat{\Omega}_{21}$ de type $k \times 1$ et $\hat{\Omega}_{22}$ de type $k \times k$.

On définit les matrices suivantes :

$$\hat{\Delta} = \hat{\Sigma} + \hat{\Lambda} = \begin{pmatrix} \hat{\Delta}_{11} & \hat{\Delta}_{12} \\ \hat{\Delta}_{21} & \hat{\Delta}_{22} \end{pmatrix} \tag{6.10}$$

$$\hat{Z} = \hat{\Delta}_{21} - \hat{\Delta}_{22}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}$$

$$\hat{y}_t^* = y_t - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\nu}_t$$

$$D = \begin{pmatrix} 0 \\ I_k \end{pmatrix}$$

où D est une matrice de type $(k+1) \times k$, 0 une matrice de type $1 \times k$, I_k de type $k \times k$, et $\hat{\nu}_t$ est le vecteur formé par tous les $\hat{\nu}_{it}$.

2. L'estimateur modifié FM (Fully Modified) de β est défini par :

$$\hat{\beta}^{FM} = \left(W'W\right)^{-1}\left(W'\hat{y}^* - nD\hat{Z}\right), \tag{6.11}$$

$$\text{où } \hat{y}^* = \begin{pmatrix} \hat{y}_1^* \\ \hat{y}_2^* \\ \vdots \\ \hat{y}_n^* \end{pmatrix}, \quad W = (\tau_n, X) \quad \text{et} \quad \tau_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{6.12}$$

Un estimateur convergent de la matrice de variance et de covariance de l'estimateur $\hat{\beta}^{FM}$ est fourni par la formule :

$$\hat{\Sigma}_{\hat{\beta}^{FM}} = \hat{\omega}_{11.2}\left(W'W\right)^{-1}, \tag{6.13}$$

$$\text{où } \hat{\omega}_{11.2} = \hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21} \tag{6.14}$$

5 Modèles *VAR* et cointégration

Une fois qu'on a vérifié la cointégration par la méthode d'Engle et Granger, et qu'on a estimé un vecteur de cointégration par MCO ou avec un estimateur Fully Modified, la relation estimée ne caractérise pas l'entièreté du lien dynamique entre les variables concernées. Elle ne mesure que la relation à long terme entre les variables et ne décrit pas toutes leurs interactions à court terme. En particulier, la relation de cointégration ne prétend pas expliquer exhaustivement toute l'évolution dynamique de l'une des variables en fonction de l'évolution dynamique des autres. D'ailleurs, son terme d'erreur doit seulement être stationnaire : il peut donc être autocorrélé. C'est bien la preuve que ce terme d'erreur peut contenir des fluctuations systématiques de la « variable de gauche » qui n'ont pas été expliquées par les fluctuations des « variables de droite ».

5.1 Définition des modèles *VAR*

Un groupe de variables aléatoires temporelles est généré par un modèle VAR si chacune de ses variables est une fonction linéaire de ses propres valeurs passées et des valeurs passées des autres variables du groupe, à laquelle s'ajoute un choc aléatoire de type bruit blanc.



Exemple

Deux variables aléatoires temporelles (processus stochastiques) X_{1t} et X_{2t} sont générées par un modèle *VAR* à deux retards, donc par un *VAR*(2), s'il existe des coefficients β_{ij} et γ_{ij}, ainsi que des variables de type bruit blanc u_{1t} et u_{2t} tels que :

$$X_{1t} = \mu_1 + \rho_{11}X_{1t-1} + \rho_{12}X_{1t-2} + \rho_{21}X_{2t-1} + \rho_{22}X_{2t-2} + u_{1t}$$
$$X_{2t} = \mu_2 + \gamma_{11}X_{1t-1} + \gamma_{12}X_{1t-2} + \gamma_{21}X_{2t-1} + \gamma_{22}X_{2t-2} + u_{2t}$$

Cela se représente encore par le système suivant :

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \rho_{11} & \rho_{12} \\ \gamma_{11} & \gamma_{12} \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} \rho_{21} & \rho_{22} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} X_{1t-2} \\ X_{2t-2} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

ou, si l'on intègre l'opérateur de retard, de la façon suivante :

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \rho_{11} & \rho_{12} \\ \gamma_{11} & \gamma_{12} \end{pmatrix} \begin{pmatrix} LX_{1t} \\ LX_{2t} \end{pmatrix} + \begin{pmatrix} \rho_{21} & \rho_{22} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} L^2X_{1t} \\ L^2X_{2t} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

ou encore :

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \rho_{11} & \rho_{12} \\ \gamma_{11} & \gamma_{12} \end{pmatrix} L \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} + \begin{pmatrix} \rho_{21} & \rho_{22} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} L^2 \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

ce qui implique :

$$\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \rho_{11} & \rho_{12} \\ \gamma_{11} & \gamma_{12} \end{pmatrix} L - \begin{pmatrix} \rho_{21} & \rho_{22} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} L^2 \right) \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

On peut alors fournir une définition générale précise d'un modèle *VAR*.

Un processus stochastique multivarié X à n composantes est généré par un modèle *VAR*(p) s'il existe un vecteur μ, des matrices Φ_l de type $n \times n$, et un processus stochastique multivarié u, dont chaque composante est un bruit blanc, tels que :

$$X_t = \mu + \sum_{l=1}^{p} \Phi_l X_{t-l} + u_t \qquad (6.15)$$

et donc tels que :

$$\left(I_n - \sum_{l=1}^{p} \Phi_l L^l \right) X_t = \mu + u_t \qquad (6.16)$$

ou encore tels que :

$$\Phi(L)X_t = \mu + u_t \qquad (6.17)$$

où $\Phi(L)$ est un polynôme matriciel en l'opérateur de retard L, défini par :

$$\Phi(L) = \left(I_n - \sum_{l=1}^{p} \Phi_l L^l \right) \qquad (6.18)$$

Exemple (suite)

Dans l'exemple précédent où $n = 2$, on a :

$$X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}, \; \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \; \Phi_1 = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \gamma_{11} & \gamma_{12} \end{pmatrix}, \; \Phi_2 = \begin{pmatrix} \rho_{21} & \rho_{22} \\ \gamma_{21} & \gamma_{22} \end{pmatrix},$$

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \text{ et } \Phi(L) = \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \rho_{11} & \rho_{12} \\ \gamma_{11} & \gamma_{12} \end{pmatrix} L - \begin{pmatrix} \rho_{21} & \rho_{22} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} L^2 \right)$$

5.2 Propriétés dynamiques d'un modèle *VAR*

L'équation caractéristique du modèle $VAR(p)$ est $0 = |\Phi(L)| = \left| I_n - \sum_{l=1}^{p} \Phi_l L^l \right|$, où la notation $|.|$ représente le déterminant. Si les racines de l'équation $0 = |\Phi(L)|$ sont en dehors du cercle unité, on montre que le polynôme $\Phi(L)$ est inversible, et donc qu'il existe une infinité de matrices C_l de type $n \times n$ telles que X_t peut aussi s'écrire comme une moyenne mobile infinie des valeurs courantes et passées du choc aléatoire multivarié u :

$$X_t = \tilde{\mu} + C_0 u_t + C_1 u_{t-1} + C_2 u_{t-2} + \cdots \tag{6.19}$$

et donc telles que :

$$\begin{aligned} X_t &= \tilde{\mu} + \sum_{l=0}^{\infty} C_l u_{t-l} \\ &= \tilde{\mu} + \sum_{l=0}^{\infty} C_l L^l u_t = \tilde{\mu} + C(L) u_t \end{aligned}$$

où $C(L) = \sum_{l=0}^{\infty} C_l L^l . X_t = \left(\Phi(L)\right)^{-1} (\mu + u_t) = \tilde{\mu} + C(L) u_t$ et donc $\left(\Phi(1)\right)^{-1} \mu = \tilde{\mu}$ et $\left(\Phi(L)\right)^{-1} = C(L)$.

L'équation caractéristique ($0 = |\Phi(L)|$) du système *VAR* admet une ou plusieurs racines unitaires si $0 = |\Phi(1)|$ et donc si la matrice $\Phi(1) = \left(I_n - \sum_{l=1}^{p} \Phi_l \right)$ est singulière.

5.3 Tests de la qualité de la spécification

Après avoir estimé un modèle *VAR*, et avant de tester des hypothèses sur les coefficients estimés, il faut vérifier si le modèle représente de manière appropriée la dynamique des variables modélisées.

Il faut d'abord calculer des tests « portmanteau » séparément pour chaque équation du *VAR* pour vérifier l'hypothèse d'absence d'autocorrélation du terme d'erreur de chacune. Ensuite, il faut réaliser un test « portmanteau » vectoriel pour vérifier en une passe l'hypothèse d'absence d'autocorrélation des termes d'erreur de toutes les équations du *VAR*.

Test « portmanteau » de Ljung et Box [LJU 1978]

Il s'agit d'une version du test de Box et Pierce (1970), corrigée par les degrés de liberté. Ce test est valable uniquement dans une équation sans autres variables explicatives que

des retards de la variable dépendante. Il permet donc de tester l'autocorrélation du terme d'erreur d'une équation particulière d'un système *VAR*. Il est calculé d'après la formule suivante :

$$Q^*(s) = N(N+2)\sum_{j=1}^{s}\frac{\left(r_j^*\right)^2}{N-j} \tag{6.20}$$

où les coefficients d'autocorrélation empirique des résidus de l'équation sont :

$$r_j^* = \frac{\sum_{t=j+1}^{N}(e_t)\left(e_{t-j}\right)}{\sqrt{\sum_{t=j+1}^{N}(e_t)^2}} \tag{6.21}$$

Le test $Q^*(s)$ est de type multiplicateur de Lagrange ; il s'agit donc d'un test LM. Sous l'hypothèse H_0 que le terme d'erreur vrai (non observable) de l'équation n'est pas autocorrélé, le test $Q^*(s)$ est distribué asymptotiquement selon une Chi-2 à $s-p$ degrés de liberté. On rejette donc l'hypothèse d'absence d'autocorrélation du terme d'erreur de l'équation si $Q^*(s)$ prend une valeur réalisée supérieure à la valeur critique d'une Chi-2 à $s-p$ degrés de liberté, au seuil de signification choisi (en général 5 %).

Test « portmanteau » de Doornik et Hendry [DOO 1997]

Il s'agit également d'une version du test de Box et Pierce [BOX 1970], mais que l'on corrige par les degrés de liberté en calculant différemment les coefficients d'autocorrélation empiriques des résidus. Il est calculé d'après la formule suivante :

$$LB(s) = N(N+2)\sum_{j=1}^{s}\frac{r_j^2}{N-j} \tag{6.22}$$

où les coefficients d'autocorrélation empirique des résidus de l'équation sont :

$$r_j = \frac{\sum_{t=j+1}^{N}(e_t-\bar{e}_0)\left(e_{t-j}-\bar{e}_j\right)}{\sqrt{\sum_{t=j+1}^{N}(e_t-\bar{e}_0)\sum\limits_{t=j+1}^{N}\left(e_{t-j}-\bar{e}_j\right)}} \tag{6.23}$$

avec $\bar{e}_0 = \dfrac{\sum_{t=j+1}^{N}e_t}{N-j}$ et $\bar{e}_j = \dfrac{\sum_{t=j+1}^{N}e_{t-j}}{N-j}$

Le test $LB(s)$ est aussi de type multiplicateur de Lagrange ; il s'agit donc d'un test *LM*. Sous l'hypothèse H_0 que le terme d'erreur vrai (non observable) de l'équation n'est pas autocorrélé, le test $LB(s)$ est distribué asymptotiquement selon une Chi-2 à $s-p$ degrés de liberté. On rejette donc l'hypothèse d'absence d'autocorrélation du terme d'erreur de l'équation si $LB(s)$ prend une valeur réalisée supérieure à la valeur critique d'une Chi-2 à $s-p$ degrés de liberté, au seuil de signification choisi (en général 5 %).

Test vectoriel « portmanteau »

Ce test est l'équivalent multivarié du test « portmanteau » développé pour une seule équation. Le test vectoriel $PB(s)$ est asymptotique et valable uniquement pour un *VAR*. Il utilise une correction pour de petits échantillons. En notant c_t le vecteur $w \times 1$ des résidus

estimés de toutes les équations du *VAR* à la période t, on définit les matrices suivantes, pour tout $r, s > 0$:

$$\hat{C}_{rs} = \frac{\sum_{t=1}^{N} c_{t-r}c'_{t-s}}{N} \tag{6.24}$$

où $c_{t-i} = 0$ pour $t - i < 1$. Le test vectoriel « portmanteau » est alors défini de la manière suivante :

$$LB(s) = N^2 \sum_{j=1}^{s} \frac{tr\left(\hat{C}'_{0j}\hat{C}_{00}^{-1}\hat{C}_{0j}\hat{C}_{00}^{-1}\right)}{N-j} \tag{6.25}$$

Il est proposé par Hosking [HOS 1980] et Lütkepohl [LUT 1991] et est implémenté par exemple dans PcFiml par Doornik et Hendry [DOO 1997]. Sous l'hypothèse H_0 d'absence d'autocorrélation des résidus de toutes les équations du $VAR(p)$, $LB(s)$ est distribué asymptotiquement selon une Chi-2 à $w^2(s-p)$ degrés de liberté. On rejette donc l'hypothèse d'absence d'autocorrélation des termes d'erreur des w équations du $VAR(p)$ si $LB(s)$ prend une valeur réalisée supérieure à la valeur critique d'une Chi-2 à $w^2(s-p)$ degrés de liberté, au seuil de signification choisi (en général 5 %). Le test $LB(s)$ est une version avec correction du test vectoriel de Box et Pierce qui est :

$$PB(s) = N \sum_{j=1}^{s} tr\left(\hat{C}'_{0j}\hat{C}_{00}^{-1}\hat{C}_{0j}\hat{C}_{00}^{-1}\right) \tag{6.26}$$

Autres tests

Pour tester l'hypothèse d'absence d'autocorrélation du terme d'erreur d'une équation du *VAR*, on dispose du test LM habituel d'autocorrélation des résidus. Il repose sur une régression auxiliaire des résidus sur les « variables de droite » de l'équation originale et sur les valeurs retardées de ces résidus à partir d'une période $t-r$ jusqu'à une période $t-s$. Sous l'hypothèse nulle, le test est distribué selon une Chi-2 à $s-r+1$ degrés de liberté. Une valeur trop élevée du test implique donc le rejet de l'hypothèse d'absence d'autocorrélation. Ce test est valable avec des endogènes retardées, contrairement au test de Durbin et Watson.

Il est possible de généraliser l'approche précédente en testant globalement l'absence d'autocorrélation des termes d'erreur de toutes les équations du système simultanément. Ce test est basé sur un système auxiliaire, avec des valeurs retardées de résidus à partir d'une période $t-r$ jusqu'à une période $t-s$. Sous l'hypothèse nulle, la version *LM* de ce test est distribuée selon une Chi-2 à $(s-r+1)n^2$ degrés de liberté. Une approximation F du test en question donne de meilleures propriétés sur de petits échantillons.

On peut utiliser un test de White habituel pour vérifier l'hypothèse d'absence d'hétéroscédasticité du terme d'erreur de chaque équation.

5.4 Modèles *VAR* et propriété de cointégration

Avec ou sans racines unitaires, on montre que $\Phi(L)$ peut se décomposer de la manière suivante :

$$\Phi(L) = \Phi(1)L + (1-L)\left(I_n - \sum_{l=1}^{p-1} A_l L^l\right) \tag{6.27}$$

où

$$A_l = -\sum_{j=l+1}^{p} \Phi_l \quad \forall l = 1, 2, \dots p-1 \tag{6.28}$$

Par conséquent, le système $VAR(p)$, dont la forme originale est $\Phi(L)X_t = \mu + u_t$, peut encore être reparamétré de la manière suivante :

$$\Delta X_t = \mu - \Phi(1)X_{t-1} + \sum_{l=1}^{p-1} A_l \Delta X_{t-l} + u_t \tag{6.29}$$

Ce reparamétrage est particulièrement utile si des racines unitaires existent, les variables X_i étant intégrées du même ordre (1). On montre alors que les variables de X sont cointégrées et que r vecteurs de cointégration linéairement indépendants existent **si** le rang de $\Phi(1)$ est égal à r ($< n$) **et s'**il existe des matrices α et β, de type $n \times r$ et de rang r, où $0 < r \leqslant n$, telles que $\Phi(1) = -\alpha\beta'$. Le système $VAR(p)$ peut alors être représenté sous forme de mécanisme à correction d'erreur, de la manière suivante :

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + \sum_{l=1}^{p-1} A_l \Delta X_{t-l} + u_t \tag{6.30}$$

La matrice β est la matrice de type $n \times r$ dont chaque colonne est un vecteur de cointégration différent. Le vecteur $\beta' X_{t-1}$, qui a r éléments, est le vecteur des r expressions cointégrées qui mesurent chacune un écart par rapport à une relation d'équilibre différente, à la période $t-1$. Si l'on explicite les matrices, la forme à correction d'erreur précédente est :

$$\begin{pmatrix} X_{1t} - X_{1t-1} \\ X_{2t} - X_{2t-1} \\ \vdots \\ X_{nt} - X_{nt-1} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1r} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2r} \\ \dots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nr} \end{bmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1r} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nr} \end{pmatrix}' \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \\ \vdots \\ X_{nt-1} \end{pmatrix}$$

$$+ \sum_{l=1}^{p-1} \begin{pmatrix} a_{11l} & a_{12l} & \cdots & a_{1nl} \\ a_{21l} & a_{22l} & \cdots & a_{2nl} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1l} & a_{n2l} & \cdots & a_{nnl} \end{pmatrix} \begin{pmatrix} X_{1t-l} - X_{1t-l-1} \\ X_{2t-l} - X_{2t-l-1} \\ \vdots \\ X_{nt-l} - X_{nt-l-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ \\ u_{nt} \end{pmatrix}$$

et donc :

$$\begin{pmatrix} X_{1t} - X_{1t-1} \\ X_{2t} - X_{2t-1} \\ \vdots \\ X_{nt} - X_{nt-1} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1r} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2r} \\ \dots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nr} \end{bmatrix} \begin{pmatrix} \beta_{11}X_{1t-1} + \beta_{21}X_{2t-1} + \cdots + \beta_{n1}X_{nt-1} \\ \beta_{12}X_{1t-1} + \beta_{22}X_{2t-1} + \cdots + \beta_{n2}X_{nt-1} \\ \vdots \\ \beta_{1r}X_{1t-1} + \beta_{2r}X_{2t-1} + \cdots + \beta_{nr}X_{nt-1} \end{pmatrix}$$

$$+ \sum_{l=1}^{p-1} \begin{pmatrix} a_{11l} & a_{12l} & \cdots & a_{1nl} \\ a_{21l} & a_{22l} & \cdots & a_{2nl} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1l} & a_{n2l} & \cdots & a_{nnl} \end{pmatrix} \begin{pmatrix} X_{1t-l} - X_{1t-l-1} \\ X_{2t-l} - X_{2t-l-1} \\ \vdots \\ X_{nt-l} - X_{nt-l-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{nt} \end{pmatrix}$$

Le $VAR(p)$ cointégré à correction d'erreur est le système d'équations suivantes pour les n variables :

$$\begin{aligned}
X_{1t} - X_{1t-1} = {} & \mu_1 + \alpha_{11}\left(\beta_{11}X_{1t-1} + \beta_{21}X_{2t-1} + \cdots + \beta_{n1}X_{nt-1}\right) \\
& + \alpha_{12}\left(\beta_{12}X_{1t-1} + \beta_{22}X_{2t-1} + \cdots + \beta_{n2}X_{nt-1}\right) + \cdots \\
& + \alpha_{1r}\left(\beta_{1r}X_{1t-1} + \beta_{2r}X_{2t-1} + \cdots + \beta_{nr}X_{nt-1}\right) + \sum_{l=1}^{p-1} a_{11l}\left(X_{1t-l} - X_{1t-l-1}\right) \\
& + \sum_{l=1}^{p-1} a_{12l}\left(X_{2t-l} - X_{2t-l-1}\right) + \cdots + \sum_{l=1}^{p-1} a_{1nl}\left(X_{nt-l} - X_{nt-l-1}\right) + u_{1t} \\
X_{2t} - X_{2t-1} = {} & \mu_2 + \alpha_{21}\left(\beta_{11}X_{1t-1} + \beta_{21}X_{2t-1} + \cdots + \beta_{n1}X_{nt-1}\right) \\
& + \alpha_{22}\left(\beta_{12}X_{1t-1} + \beta_{22}X_{2t-1} + \cdots + \beta_{n2}X_{nt-1}\right) + \cdots \\
& + \alpha_{2r}\left(\beta_{1r}X_{1t-1} + \beta_{2r}X_{2t-1} + \cdots + \beta_{nr}X_{nt-1}\right) + \sum_{l=1}^{p-1} a_{21l}\left(X_{1t-l} - X_{1t-l-1}\right) \\
& + \sum_{l=1}^{p-1} a_{22l}\left(X_{2t-l} - X_{2t-l-1}\right) + \cdots + \sum_{l=1}^{p-1} a_{2nl}\left(X_{nt-l} - X_{nt-l-1}\right) + u_{2t} \\
& \vdots \\
X_{nt} - X_{nt-1} = {} & \mu_n + \alpha_{n1}\left(\beta_{11}X_{1t-1} + \beta_{21}X_{2t-1} + \cdots + \beta_{n1}X_{nt-1}\right) \\
& + \alpha_{n2}\left(\beta_{12}X_{1t-1} + \beta_{22}X_{2t-1} + \cdots + \beta_{n2}X_{nt-1}\right) + \cdots \\
& + \alpha_{nr}\left(\beta_{1r}X_{1t-1} + \beta_{2r}X_{2t-1} + \cdots + \beta_{nr}X_{nt-1}\right) \\
& + \sum_{l=1}^{p-1} a_{n1l}\left(X_{1t-l} - X_{1t-l-1}\right) + \sum_{l=1}^{p-1} a_{n2l}\left(X_{2t-l} - X_{2t-l-1}\right) + \cdots \\
& + \sum_{l=1}^{p-1} a_{nnl}\left(X_{nt-l} - X_{nt-l-1}\right) + u_{nt}
\end{aligned}$$

Avec la notation matricielle, ce long système d'équations s'écrit simplement $\Delta X_t = \mu + \alpha\beta' X_{t-1} + \sum_{l=1}^{p-1} A_l \Delta X_{t-l} + u_t$. Toutefois, l'explicitation des matrices facilite l'interprétation de leurs coefficients :

- La matrice β est telle que chacune des r colonnes est l'un des r vecteurs de cointégration linéairement indépendants.
- Le produit matriciel $\beta' X_{t-1}$ est un vecteur dont chacun des r éléments est l'une des r relations cointégrées.
- Chaque coefficient α_{ij} de la matrice A représente le taux de réaction de la i^e variable à l'écart précédent par rapport à la j^e relation d'équilibre. Il s'agit donc bien du taux de correction de la variable i suite à l'observation de l'erreur [(1)] précédente, ou du déséquilibre précédent, par rapport à la j^e relation d'équilibre. D'où le nom de mécanisme à correction d'erreur.
- Chaque coefficient a_{ipl} d'une matrice A_l représente le taux de réaction de la i^e variable à la variation passée, l périodes auparavant, de la p^e variable.

1. Le mot « erreur » désigne ici une expression cointégrée qui est un écart par rapport à un équilibre. Il n'a pas le même sens que dans l'expression « terme d'erreur », où il est synonyme de « perturbation ».



Pour des variables intégrées, cointégration et existence d'un modèle à correction d'erreur représentent le même phénomène. En somme, le fait que les variables évoluent selon un mécanisme à correction d'erreur permet la cointégration. En effet, dans un tel mécanisme, tout écart par rapport à la relation d'équilibre suscite, à la période suivante, une réaction visant à réduire cet écart. De cette manière, les variables restent liées à long terme par la relation d'équilibre : il y a cointégration entre ces variables. Ce raisonnement permet de comprendre le théorème de représentation de Granger, d'après lequel des variables sont cointégrées si et seulement si leur évolution conjointe est caractérisée par un modèle à correction d'erreur.

On suppose que toutes les variables de X_t sont intégrées d'ordre 1. On montre que la cointégration (rang de $\Phi(1) = r < n$, $\Phi(1) = \alpha\beta'$, $\beta' X_t \sim I(0)$) implique que :

$$X_t = X_0 + C(1)\left(\mu t + \sum_{j=1}^{t} u_j\right) + C^*(L)U_t \tag{6.31}$$

où

- $C(L)$ est le polynôme en l'opérateur de retard L de la représentation de Wold de DX_t :

$$\Delta X_t = C(L)\,(\mu + u_t) = \tilde{\mu} + C(L)u_t \quad \text{où } \tilde{\mu} = C(1)\mu$$

- $C^*(L)$ est défini par la décomposition suivante de $C(L)$:

$$C(L) = C(1) + C^*(L)(1 - L)$$

- $C(1) = \beta_\perp \left(\alpha'_\perp \Psi(1)\beta_\perp\right)^{-1} \alpha'_\perp$ où $\alpha_\perp$ et $\beta_\perp$ sont des matrices de type $n \times (n - r)$ et de rang $n - r$, telles que $\alpha_{\perp'}\alpha = 0$ et $\beta_{\perp'}\beta = 0$, et où $\Psi(L) = -\sum_{j=0}^{p-1} A_j L^j$ et donc $\Psi(1) = -\sum_{j=0}^{p-1} A_j$.

Dans cette égalité, on observe que les variables composant X_t ont une tendance stochastique $\sum_{j=1}^{t} u_j$, ce qui est logique puisque X_t est $I(1)$, mais aussi qu'elles ont une tendance déterministe t, sauf si $C(1)\mu = 0$. On en tire les remarques suivantes :

- Si les variables X_{it} composant X_t ont une racine unitaire avec dérive, on sait qu'elles ont effectivement une tendance déterministe, en plus de la tendance stochastique (voir chapitre 5). Il est alors logique de laisser μ sans restriction, de ne pas imposer $C(1)\mu = 0$.
- Si les variables X_{it} composant X_t ont une racine unitaire sans dérive, on sait qu'elles n'ont qu'une tendance stochastique et pas de tendance déterministe (voir chapitre 5). Il est alors logique d'imposer que $C(1)\mu = 0$.

En pratique, si l'on étudie n variables X_{it} qui n'ont pas visuellement de tendance croissante ou décroissante, il est approprié d'imposer la contrainte $C(1)\mu = 0$. On montre que cela revient à imposer le vecteur μ comme une combinaison linéaire des colonnes de α.

6 Tests de cointégration de Johansen

6.1 Principe

Les tests de Johansen permettent de vérifier des hypothèses sur le nombre de vecteurs de cointégration dans un système $VAR(p)$ reliant des variables qui sont toutes intégrées du même ordre.

On part d'un vecteur X_t (rassemblant n variables) dont l'évolution est décrite par un modèle $VAR(p)$ du type :

$$X_t = \mu + \sum_{l=1}^{p} \Phi_l X_{t-l} \left(+ \sum_{i=1}^{s-1} w_i Q_{it} \right) + u_t \tag{6.32}$$

On peut le reparamétrer ainsi :

$$\Delta X_t = \mu - \Phi(1) X_{t-1} + \sum_{l=1}^{p-1} A_l \Delta X_{t-l} + \left(\sum_{i=1}^{s-1} w_i Q_{it} \right) + u_t \tag{6.33}$$

avec :

$$A_l = -\sum_{j=l+1}^{p} \Phi_l \quad \forall l = 1, 2, \ldots p-1 \tag{6.34}$$

$$\Phi(1) = I_n - \sum_{j=l+1}^{p} \Phi_l \tag{6.35}$$

Les variables Q_{it} sont d'éventuelles variables indicatrices (ou dummy), saisonnières (voir chapitre 3), s représentant la fréquence des données (12 pour des données mensuelles par exemple). On peut rarement ajouter une tendance linéaire t. On suppose que $u_t \sim N_n(0, \Sigma)$ et que les u_t de périodes différentes sont indépendants. On suppose aussi que $X_t \sim I(1)$, donc que toutes les variables sont intégrées d'ordre 1. Sous l'hypothèse de normalité de u, la méthode de Johansen permet de tester l'hypothèse de l'existence d'un effectif r quelconque de relations cointégrées, allant de 0 à $n-1$.

Johansen montre que les estimateurs de maximum de vraisemblance des paramètres d'un modèle *VAR* cointégré sont fonction de quelques matrices assemblées à partir des résultats de l'estimation de régressions auxiliaires par MCO. Pour appliquer la méthode de Johansen, on réalise donc d'abord l'estimation par MCO du système d'équations linéaires suivant :

$$\Delta X_{t-1} = \hat{C}_0 + \hat{C}_{01} \Delta X_{t-1} + \cdots + \hat{C}_{0p-1} \Delta X_{t-p-1} + \left(\sum_{i=1}^{s-1} \hat{w}_i Q_{it} \right) + R_{0t} \tag{6.36}$$

où C_0 est une matrice de type $m \times 1$, chaque C_{0i} est de type $m \times m$ et R_{0t} est un vecteur de m résidus au temps t.

On réalise également l'estimation par MCO du système d'équations linéaires suivant :

$$X_{t-1} = \hat{C}_1 + \hat{C}_{11} \Delta X_{t-1} + \cdots + \hat{C}_{1p-1} \Delta X_{t-p-1} + \left(\sum_{i=1}^{s-1} \hat{w}_i Q_{it} \right) + R_{1t} \tag{6.37}$$

où C_1 est une matrice de type $n \times 1$, chaque C_{1i} est de type $n \times n$ et R_{1t} est un vecteur de m résidus au temps t.

On utilise alors les vecteurs R_{0t} et R_{1t} de résidus calculés pour former les matrices S_{01}, S_{00}, S_{10} et S_{11}, de type $N \times N$, de la manière suivante :

$$S_{ij} = \left(\frac{1}{N} \right) \sum_{t=1}^{N} R_{it} R'_{jt} \tag{6.38}$$

où N est l'effectif des observations disponibles (les variables X_i sont observées sur N périodes).

Pourquoi tous ces calculs préliminaires ? On montre aisément que la fonction de vraisemblance du système $VAR(p)$ avec $\Phi(1) = -\alpha\beta'$ est maximisée si l'on choisit les estimateurs :

$$\hat{\Sigma}^{MV} = S_{00} - S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta' S_{10} \quad (6.39)$$

$$\hat{\alpha}^{MV} = S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1} \quad (6.40)$$

À ce stade, ces estimateurs sont fonction de β, qui est lui-même une matrice de paramètres inconnus qu'il faut estimer. Si l'on a un estimateur de maximum de vraisemblance pour β, il suffit de le remplacer dans les formules précédentes pour disposer des estimateurs de maximum de vraisemblance de Σ et α.

La fonction de vraisemblance du système $VAR(p)$ avec $\Phi(1) = \alpha\beta$, où α et β ont r colonnes et un rang r, est maximisée si l'on choisit $\hat{\beta}^{MV} = \left(\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_r\right)$, où $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n$ sont les vecteurs propres associés aux n valeurs propres $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_n$ dont le vecteur $\hat{\lambda} = \left(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_n\right)'$ est solution du système d'équations suivant :

$$\left|\hat{\lambda} S_{11} - S_{10} S_{00}^{-1} S_{01}\right| = 0 \quad (6.41)$$

Elles sont classées par ordre décroissant : $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. Les vecteurs propres $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n$ sont normalisés de manière à ce que leur matrice $V = (\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n)$, de type $n \times n$, vérifie :

$$I_n = V' S_{11} V \quad (6.42)$$

Les r vecteurs cointégrants estimés, donc les r colonnes de $\hat{\beta}^{MV}$, sont ainsi les r premiers vecteurs propres (les vecteurs associés aux r plus grandes valeurs propres parmi les n valeurs propres), parmi les n vecteurs propres existants. (Les r vecteurs de cointégration sont les r vecteurs propres les plus significatifs parmi les m vecteurs propres.)

Ces calculs d'estimateurs sont bien sûr conditionnels au choix d'un rang r. Lequel choisir ? Pour déterminer le rang r de cointégration, Johansen propose deux tests :

a) Le test de la trace, $LR(r|n) = -N \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i)$, permet de vérifier l'hypothèse H_0 : rang de $\beta = r$ (r vecteurs cointégrants linéairement indépendants) contre l'hypothèse H_1 : rang de $\beta = n$ (n vecteurs cointégrants linéairement indépendants). Sous l'hypothèse H_0, ce test a une distribution particulière, tabulée par Johansen.

b) Le test de la plus grande valeur propre, $LR(r|n) = -N \ln(1 - \hat{\lambda}_{r+1})$, permet de vérifier l'hypothèse H_0 : rang de $\beta = r$ (r vecteurs cointégrants linéairement indépendants) contre l'hypothèse H_1 : rang de $\beta = r + 1$ ($r + 1$ vecteurs cointégrants linéairement indépendants). Sous l'hypothèse H_0, ce test a une distribution particulière tabulée par Johansen.

Ces deux outils sont valables lorsque μ est laissé sans restriction, ce qui permet à X_t d'avoir une tendance linéaire. Ils se déclinent en d'autres versions :

- On peut imposer *a priori* que $\alpha_{\perp}\mu = 0$, c'est-à-dire qu'il existe un vecteur β_0 de type $r \times 1$ tel que $\mu = \alpha\beta_0$; sous les mêmes hypothèses H_0 précédentes, les tests ont des distributions particulières tabulées par Johansen.
- On peut également imposer que $\mu = 0$.
- On peut ajouter une tendance linéaire dans le modèle et en contraindre le coefficient pour éviter que les variables de X_t aient une tendance quadratique.

Remarque

Le test de la trace a souvent des distorsions de niveau ou de seuil : l'hypothèse H_0 d'absence de cointégration ($r = 0$), ou bien l'hypothèse H_0 d'une autre valeur de r, est souvent erronément rejetée, alors qu'elle est vraie, en faveur de l'hypothèse H_1 d'un r supérieur. Tel qu'il est programmé dans TSP, le test de la trace incorpore une correction de petits échantillons proposée par Gregory [GRE 1994].

Avec TSP, les probabilités critiques (p-values) des tests de Johansen sont obtenues par interpolation à partir des tables d'Osterwald-Lenum [OST 1992].

6.2 Identification des vecteurs de cointégration

La technique d'estimation de Johansen permet donc d'obtenir r vecteurs de cointégration estimés, mais qui ne sont pas identifiés puisque toute combinaison linéaire de ces vecteurs estimés peut aussi être considérée comme un vecteur de cointégration. Pour les identifier, il faut imposer des *restrictions de normalisation* et des *restrictions supplémentaires d'identification*. Il s'agit de choisir des vecteurs $\tilde{\hat{\beta}}_1, \tilde{\hat{\beta}}_2 \dots \tilde{\hat{\beta}}_r$ particuliers (qui sont les colonnes d'une matrice $\tilde{\hat{\beta}}$) parmi l'infinité de vecteurs définis comme des combinaisons linéaires des vecteurs estimés $\hat{\beta}_1, \hat{\beta}_2 \dots \hat{\beta}_r$ (qui sont les colonnes d'une matrice $\hat{\beta}$).

Les restrictions de normalisation sont :

$$\tilde{\hat{\beta}}_{ii} = 1 \quad \forall i = 1, \dots, r \tag{6.43}$$

Les restrictions supplémentaires d'identification peuvent être imposées, en même temps que les précédentes, de la manière suivante[1] :

$$\tilde{\hat{\beta}} = \hat{\beta} \left(c' \hat{\beta} \right)^{-1} \tag{6.44}$$

où c est une matrice $n \times r$ de coefficients dont les valeurs sont fixées arbitrairement en fonction des restrictions qu'on veut imposer. Par exemple, si l'on veut imposer que les r premières variables n'interviennent chacune que dans une seule des r relations de cointégration, on choisit[2] :

$$c = \begin{pmatrix} I_r \\ O \end{pmatrix} \tag{6.45}$$

1. On impose en fait les restrictions $c'' \tilde{\hat{\beta}} = I_r$.

2. I_r est la matrice identité à r lignes et r colonnes : $I_r = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & \ddots & 0 & 1 \end{pmatrix}$, tandis que O est une matrice $(n - r) \times r$ qui ne contient que des 0 : $O = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$.



On obtient une matrice de cette forme :

$$\tilde{\hat{\beta}} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \tilde{\hat{\beta}}_{(r+1)1} & \tilde{\hat{\beta}}_{(r+1)2} & \cdots & \tilde{\hat{\beta}}_{(r+1)(r-1)} & \tilde{\hat{\beta}}_{(r+1)r} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \tilde{\hat{\beta}}_{n1} & \tilde{\hat{\beta}}_{n2} & \cdots & \tilde{\hat{\beta}}_{n(r-1)} & \tilde{\hat{\beta}}_{nr} \end{pmatrix} \qquad (6.46)$$

Dans ces conditions, la méthode d'Engle et Granger, si elle n'est appliquée que sur un sous-ensemble de $n+1-r$ variables [1] parmi les n variables intégrées, est une technique qui permet d'emblée d'estimer isolément l'une des r relations de cointégration normalisées et restreintes.

$$\tilde{\beta} = \beta\left(c'\beta\right)^{-1} \qquad (6.47)$$

Exemple

Soient n variables $I(1)$ cointégrées et deux vecteurs de cointégration linéairement indépendants ($r = 2$) estimés :

$$\begin{pmatrix} \hat{\beta}_{11} \\ \hat{\beta}_{21} \\ \vdots \\ \hat{\beta}_{n1} \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} \hat{\beta}_{12} \\ \hat{\beta}_{22} \\ \vdots \\ \beta_{n2} \end{pmatrix}$$

Ils forment la matrice :

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} \\ \hat{\beta}_{21} & \hat{\beta}_{22} \\ \vdots & \vdots \\ \hat{\beta}_{n1} & \hat{\beta}_{n2} \end{pmatrix}$$

Si l'on définit la matrice c par...

$$c = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}$$

1. Une seule parmi les r premières et toutes les $n-r$ dernières.

... on obtient :

$$\widetilde{\hat{\beta}} = \hat{\beta}\left(c'\hat{\beta}\right)^{-1} = \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} \\ \hat{\beta}_{21} & \hat{\beta}_{22} \\ \vdots & \vdots \\ \hat{\beta}_{n1} & \hat{\beta}_{n2} \end{pmatrix} \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} \\ \hat{\beta}_{21} & \hat{\beta}_{22} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} \\ \hat{\beta}_{21} & \hat{\beta}_{22} \\ \vdots & \vdots \\ \hat{\beta}_{n1} & \hat{\beta}_{n2} \end{pmatrix} \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} \\ \hat{\beta}_{21} & \hat{\beta}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} \\ \hat{\beta}_{21} & \hat{\beta}_{22} \\ \vdots & \vdots \\ \hat{\beta}_{n1} & \hat{\beta}_{n2} \end{pmatrix} \left(\frac{1}{\hat{\beta}_{11}\hat{\beta}_{22} - \hat{\beta}_{12}\hat{\beta}_{21}}\right) \begin{pmatrix} \hat{\beta}_{22} & -\hat{\beta}_{12} \\ -\hat{\beta}_{21} & \hat{\beta}_{11} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \dfrac{\left(\hat{\beta}_{31}\hat{\beta}_{22} - \hat{\beta}_{32}\hat{\beta}_{21}\right)}{\left(\hat{\beta}_{11}\hat{\beta}_{22} - \hat{\beta}_{12}\hat{\beta}_{21}\right)} & \dfrac{\left(-\hat{\beta}_{31}\hat{\beta}_{12} + \hat{\beta}_{32}\hat{\beta}_{11}\right)}{\left(\hat{\beta}_{11}\hat{\beta}_{22} - \hat{\beta}_{12}\hat{\beta}_{21}\right)} \\ \vdots & \vdots \\ \dfrac{\left(\hat{\beta}_{n1}\hat{\beta}_{22} - \hat{\beta}_{n2}\hat{\beta}_{21}\right)}{\left(\hat{\beta}_{11}\hat{\beta}_{22} - \hat{\beta}_{12}\hat{\beta}_{21}\right)} & \dfrac{\left(-\hat{\beta}_{n1}\hat{\beta}_{12} + \hat{\beta}_{n2}\hat{\beta}_{11}\right)}{\left(\hat{\beta}_{11}\hat{\beta}_{22} - \hat{\beta}_{12}\hat{\beta}_{21}\right)} \end{pmatrix}$$

Les nouveaux vecteurs $\widetilde{\hat{\beta}}_1$ (première colonne de $\widetilde{\hat{\beta}}$) et $\widetilde{\hat{\beta}}_2$ (deuxième colonne de $\widetilde{\hat{\beta}}$) ainsi obtenus sont bien des combinaisons linéaires des vecteurs $\hat{\beta}_1$ et $\hat{\beta}_2$ initiaux :

$$\widetilde{\hat{\beta}}_1 = \frac{\hat{\beta}_{22}}{\left(\hat{\beta}_{11}\hat{\beta}_{22} - \hat{\beta}_{12}\hat{\beta}_{21}\right)}\hat{\beta}_1 - \frac{\hat{\beta}_{21}}{\left(\hat{\beta}_{11}\hat{\beta}_{22} - \hat{\beta}_{12}\hat{\beta}_{21}\right)}\hat{\beta}_2$$

7 Modèles à correction d'erreur conditionnel

Le vecteur u_t des termes d'erreur du système dynamique cointégré est supposé non autocorrélé et distribué d'après une normale, de matrice de variance et de covariance Σ c'est-à-dire : $u_t \sim N(0, \Sigma)$. Il en résulte la distribution conditionnelle [1] suivante de ΔX_t par rapport à « son passé » :

$$\Delta X_t \,|X_{t-1}, X_{t-2}, \ldots \sim N\left(\mu + \alpha\beta' X_{t-1} + A_1 \Delta X_{t-1}, \Sigma\right) \tag{6.48}$$

On partitionne le vecteur X_t en $X_t = \begin{pmatrix} Y_t \\ Z_t \end{pmatrix}$, où Y_t est un vecteur $m \times 1$ et Z_t un vecteur $(n-m) \times 1$:

$$Y_t = \begin{pmatrix} X_{1t} \\ \vdots \\ X_{mt} \end{pmatrix}, \quad Z_t = \begin{pmatrix} X_{(m+1)t} \\ X_{(m+2)t} \\ \vdots \\ X_{nt} \end{pmatrix} \tag{6.49}$$

1. Pour quelques rappels sur la notion de distribution conditionnelle, on peut se référer à tout bon manuel de statistique de base, comme par exemple celui de Patrick Roger, Probabilités, statistique et processus stochastiques, publié chez Pearson Education France dans la même collection.



On partitionne de même le vecteur μ et la matrice Σ :

$$\mu^1 = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}, \quad \mu^2 = \begin{pmatrix} \mu_{m+1} \\ \mu_{m+2} \\ \vdots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{6.50}$$

où

$$\Sigma_{11} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix}, \quad \Sigma_{12} = \begin{pmatrix} \sigma_{1(m+1)} & \sigma_{1(m+2)} & \cdots & \sigma_{1n} \\ \vdots & \vdots & \ddots & \sigma_{2n} \\ \sigma_{m(m+1)} & \sigma_{m(m2)} & \cdots & \sigma_{mn} \end{pmatrix},$$

$$\Sigma_{21} = \begin{pmatrix} \sigma_{(m+1)1} & \cdots & \sigma_{(m+1)m} \\ \sigma_{(m+1)2} & \cdots & \sigma_{(m+1)m} \\ & \ddots & \\ \sigma_{(m+1)2} & \cdots & \sigma_{(m+1)2} \end{pmatrix} \text{ et } \Sigma_{22} = \begin{pmatrix} \sigma_{(m+1)(m+1)} & \sigma_{(m+1)(m+2)} & \cdots & \sigma_{(m+1)n} \\ \sigma_{(m+2)(m+1)} & \sigma_{(m+2)(m+2)} & \cdots & \sigma_{(m+2)n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n(m+1)} & \sigma_{n(m+2)} & \cdots & \sigma_{nn} \end{pmatrix}$$

On partitionne également les matrices α et A_1 :

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

où α_1 est une matrice $m \times r$ contenant les m premières lignes de α, et α_2 est une matrice $(n-m) \times r$ contenant les $(n-m)$ dernières lignes de α.

$$A_1 = \begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix}$$

où A_{11} est une matrice $m \times n$ avec les m premières lignes de A, et A_{21} est une matrice $(n-m) \times n$ avec les $n-m$ dernières lignes de A.

L'hypothèse de normalité évoquée précédemment permet de déduire la distribution conditionnelle de ΔY_t conditionnellement à Z_t et au passé de X, et la distribution marginale de ΔZ_t :

$$\begin{gathered} \Delta Y_t \,|Z_t, X_{t-1}, X_{t-2} \ldots \\ \sim N \begin{pmatrix} \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\Delta Z_t - \mu_2) + \left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)\beta' X_{t-1} + \left(A_{11} - \Sigma_{12}\Sigma_{22}^{-1}A_{21}\right)\Delta X_{t-1} \\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix} \end{gathered} \tag{6.51}$$

$$\Delta Z_t \,|Z_{t-1}, Z_{t-2} \ldots \sim N\left(\mu_2 + \alpha_2\beta' X_{t-1} + A_{21}\Delta X_{t-1}, \Sigma_{22}\right) \tag{6.52}$$

On écrit encore :

$$\begin{aligned} \Delta Y_t &= \left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2\right) + \left(\Sigma_{12}\Sigma_{22}^{-1}\right)\Delta Z_t + \left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)\beta' X_{t-1} \\ &\quad + \left(A_{11} - \Sigma_{12}\Sigma_{22}^{-1}A_{21}\right)\Delta X_{t-1} + u_{1.2t} && (6.53) \\ \Delta Z_t &= \mu_2 + \alpha_2\beta' X_{t-1} + A_{21}\Delta X_{t-1} + u_{2t} && (6.54) \end{aligned}$$

où $u_{1.2t} = u_t^1 - \Sigma_{12}\Sigma_{22}^{-1}u_t^2$.

La première équation matricielle (6.53) est un *modèle à correction d'erreur conditionnel en forme réduite* tandis que la deuxième (6.54) est un *modèle à correction d'erreur marginal* [BOS 1992]. L'une et l'autre sont un simple reparamétrage du système *VAR* cointégré original. Elles contiennent donc la même information, mais présentée différemment.

$\left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2\right)$ et $u_{1.2t}$ sont des vecteurs $m \times 1$, $\Sigma_{12}\Sigma_{22}^{-1}$ est une matrice $m \times (n-m)$, $\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)$ est une matrice $m \times m$, $\beta' X_{t-1}$ est un vecteur $m \times 1$, $\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)\beta'$ est une matrice $m \times n$, et $\left(A_{11} - \Sigma_{12}\Sigma_{22}^{-1}A_{21}\right)$ une matrice $m \times n$.

8 Modélisation structurelle

Après avoir défini un modèle *VAR* à correction d'erreur conditionnel, il est intéressant de définir un modèle structurel faisant apparaître des interactions entre les variables d'intérêt (celles de Y) contemporaines. L'avantage d'une telle modélisation est donc que les variables contemporaines d'intérêt sont incluses dans l'équation, ce qui n'était pas le cas auparavant.

En reparamétrant le modèle à correction d'erreur conditionnel en forme réduite, on peut spécifier différents modèles structurels, en particulier un modèle structurel à équations simultanées à correction d'erreur.

On partitionne la matrice β' de la manière suivante : $\beta' = \left(\Gamma \vdots B\right)$, où Γ est supposé être une matrice non singulière $m \times m$ et B est une matrice $m \times (n-m)$, ce qui induit :

$$\beta' X_t = \Gamma Y_t + B Z_t = v_t \tag{6.55}$$

L'introduction d'une partition de la matrice β en deux matrices Γ et B permet simplement de réinterpréter les relations de cointégration dans une perspective de modèle structurel reliant des variables endogènes Y à des variables exogènes Z. Le vecteur $m \times 1 v_t$ est ainsi défini comme le vecteur des écarts aux équilibres à la période t.

Les matrices Γ et B ne sont identifiées que moyennant des hypothèses d'identification, décrites par Park [PAR 1990] et Saikkonen [SAI 1991]. Ces matrices n'étant qu'une partition de β, les hypothèses d'identification exacte de β examinées précédemment valent aussi pour Γ et B, et avec la nouvelle notation de β partitionné en Γ et B, elles peuvent s'écrire de la manière suivante [1] :

$\Gamma_{ii} = -1, i = 1 \ldots m$ pour les hypothèses de normalisation et $R_i \left(\Gamma_i \vdots B_i\right)' = 0, i = 1 \ldots m$ pour les hypothèses supplémentaires d'identification exacte.

Le vecteur $\left(\Gamma_i \vdots B_i\right)'$ est la i^e colonne de β. Chaque matrice de restrictions R_i doit être d'ordre $(r-1) \times n$ et de rang $r-1$: un vecteur de cointégration particulier est exactement identifié si, en plus de la restriction de normalisation, on lui impose $r-1$ restrictions linéairement indépendantes.

Bien entendu, on peut également imposer, en plus des restrictions de normalisation et des restrictions supplémentaires d'identification exacte, des restrictions supplémentaires de suridentification. C'est le cas ici si une ou plusieurs matrices R_i ont plus de $r-1$ lignes et sont d'un rang supérieur à $r-1$. Seules les restrictions de suridentification peuvent être testées statistiquement.

1. En l'absence de restrictions « interéquations » qui seraient une autre manière d'identifier les coefficients.



Un *modèle structurel à correction d'erreur* peut s'obtenir simplement par reparamétrage du modèle à correction d'erreur conditionnel en forme réduite : on substitue $\left(\Gamma \vdots B\right)$ à β' et l'on multiplie l'ensemble par une matrice non singulière $m \times m$ notée Γ_0 :

$$\begin{aligned}\Gamma_0 \Delta Y_t = \Gamma_0\left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2\right) + \Gamma_0\left(\Sigma_{12}\Sigma_{22}^{-1}\right)\Delta Z_t \\ + \Gamma_0\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)\left(\Gamma Y_{t-1} + BZ_{t-1}\right) \\ + \Gamma_0\left(A_{11} - \Sigma_{12}\Sigma_{22}^{-1}A_{21}\right)\Delta X_{t-1} + \Gamma_0 u_{1.2t}\end{aligned} \quad (6.56)$$

Le nouveau vecteur de termes d'erreur $\xi_t = \Gamma_0 u_{1.2t}$ a une matrice de variance et de covariance $\Xi = \Gamma_0\left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)\Gamma_0'$.
Le modèle structurel peut alors être noté de manière plus simple :

$$\Gamma_0 \Delta Y_t = b + B_0 \Delta Z_t + \Lambda\left(\Gamma Y_{t-1} + BZ_{t-1}\right) + \Gamma_1 \Delta Y_{t-1} + B_1 \Delta Z_{t-1} + \xi_t \quad (6.57)$$

moyennant les définitions suivantes :

$$b = \Gamma_0\left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2\right), \quad B_0 = \Gamma_0\left(\Sigma_{12}\Sigma_{22}^{-1}\right), \quad \Lambda = \Gamma_0\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)$$

où $\left(\Gamma_1 \vdots B_1\right) = \Gamma_0\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)$

Ces nouvelles notations sont des matrices dont les éléments sont notés ainsi :

$$\Gamma_0 = \begin{pmatrix} \gamma_{0,11} & \cdots & \gamma_{0,1m} \\ \vdots & \ddots & \vdots \\ \gamma_{0,m1} & \cdots & \gamma_{0,mm} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, \quad B_0 = \begin{pmatrix} b_{0,11} & \cdots & b_{0,1m} \\ \vdots & \ddots & \vdots \\ b_{0,m1} & \cdots & b_{0,mm} \end{pmatrix},$$

$$\Lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1m} \\ \vdots & \ddots & \vdots \\ \lambda_{m1} & \cdots & \lambda_{mm} \end{pmatrix},$$

$$\Gamma_1 = \begin{pmatrix} \gamma_{11,1} & \cdots & \gamma_{1m,1} \\ \vdots & \ddots & \vdots \\ \gamma_{m1,1} & \cdots & \gamma_{mm,1} \end{pmatrix}, \quad B_1 = \begin{pmatrix} b_{11,1} & b_{12,1} & \cdots & b_{1(n-m),1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1,1} & b_{m2,1} & \cdots & b_{m(n-m),1} \end{pmatrix}$$

Le problème de l'identification de Γ et B ayant déjà été traité, il faut encore aborder celui de l'identification des paramètres restants : $\left(\Gamma_1 \vdots B_1\right) = \Gamma_0\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)$. Il s'agit des paramètres décrivant les ajustements à court terme (Γ et B étant les paramètres des relations d'équilibre à long terme). À nouveau, il faut imposer des restrictions de normalisation et des restrictions supplémentaires d'identification, inspirées si possible par la théorie économique. Les restrictions de normalisation peuvent s'exprimer de la manière suivante :

$$\Gamma_{ii} = 1, \quad i = 1 \ldots m \quad (6.58)$$

La sélection de restrictions supplémentaires d'identification sur les paramètres à court terme est difficile, car les propositions dégagées par la théorie économique portent essentiellement sur les relations d'équilibre et expliquent généralement peu les ajustements à court terme.

Il serait donc raisonnable de commencer par des restrictions d'identification exacte théoriquement plausibles, et d'appliquer éventuellement ensuite, au vu des résultats, des restrictions supplémentaires de suridentification « testables ». On applique ici la méthodologie proposée initialement par Boswijk [BOS 1992], qui suggère de faire les restrictions d'identification exacte[1] qui contraignent la matrice Λ à être diagonale :

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{mm} \end{pmatrix} \tag{6.59}$$

Pour chacune des m équations, on impose donc $m - 1$ restrictions sur les coefficients, en l'occurrence des restrictions de nullité. Avec la restriction de normalisation, on dispose bien de m restrictions, qui permettent l'identification exacte.

Résumé

Les tests de cointégration permettent de vérifier l'existence d'une relation d'équilibre à long terme entre des variables intégrées. Des variables cointégrées suivent alors une dynamique à correction d'erreur autour de cette relation d'équilibre. La méthode de test de cointégration la plus simple est celle d'Engle et Granger, qui suppose implicitement qu'un seul vecteur cointégrant relie les variables. Même en cas de cointégration, les estimateurs de MCO des coefficients cointégrants ne suivent pas l'inférence statistique classique. Toutefois, des estimateurs modifiés la suivent. La méthode de Johansen permet de tester l'hypothèse de l'existence de plusieurs vecteurs cointégrants entre des variables intégrées, dont la dynamique est représentée par un modèle *VAR*. Un modèle *VAR* cointégré à correction d'erreur reliant un ensemble de variables intégrées permet d'obtenir un modèle *VAR* conditionnel en correction d'erreur, pour un sous-ensemble des variables dont on veut représenter la dynamique conditionnellement aux valeurs passées et courantes des autres variables. Pour ce même sous-ensemble de variables, on peut également obtenir un modèle structurel à correction d'erreur capable de représenter les relations contemporaines simultanées reliant ces variables entre elles, tout en représentant leur dépendance par rapport à leurs valeurs passées, et conditionnellement aux autres variables.

1. Cela ne revient pas ici à imposer sur la i^e relation à court terme les mêmes restrictions que celles appliquées à la i^e relation d'équilibre, alors que c'est le cas dans l'application de Boswijk.



Problèmes et exercices

Exercice 1 Assurance et test d'Engle et Granger

Énoncé

On souhaite expliquer l'évolution temporelle de la série observée du total annuel des primes d'assurance collectées par les sociétés françaises d'assurance en dommages de 1965 à 1999. Les modèles financiers suggèrent que ces primes doivent être une fonction négative du taux d'intérêt. En effet, les assureurs collectent les primes, puis les utilisent partiellement pour payer les sinistres des assurés ; entre-temps, ils les placent (en obligations...) : plus le taux d'intérêt est élevé, plus le profit financier est élevé, et donc plus faibles peuvent être les primes par rapport à la valeur espérée des sinistres à indemniser. Selon un même raisonnement, on pourrait avancer que les primes sont une fonction négative du taux de rendement boursier, ou du temps moyen entre la perception des primes et le paiement éventuel d'indemnités : plus ce temps est long, plus les placements rapportent des intérêts, et donc plus faibles peuvent être les primes exigées par les assureurs. Ce temps est appelé « longueur de branche ». Le raisonnement économique suggère aussi que les primes sont une fonction positive du taux d'inflation : en cas d'augmentation de celui-ci, les indemnités de sinistres coûtent plus cher et donc les primes exigées augmentent. Elles devraient également augmenter avec le montant observé des sinistres à indemniser et avec les dépenses de gestion des sociétés d'assurance. Elles devraient diminuer avec le niveau du capital des sociétés d'assurance : plus celui-ci est élevé, plus les sociétés d'assurance sont solvables et donc plus elles peuvent prendre le risque de diminuer les primes pour attirer de nouveaux clients. Tout cela suggère qu'à long terme, on a une relation du type :

$$\begin{aligned}\ln(\text{primes}) = \mu + \beta_2 \text{ taux d'intérêt} + \beta_3 \ln(\text{indice boursier}) + \beta_4 \ln(\text{sinistres}) \\ + \beta_5 \ln(\text{dépenses}) + \beta_6 \ln(\text{capital}) + \beta_7 \text{longueur} + \beta_8 \text{inflation} + u\end{aligned}$$

On vérifie par des tests de racine unitaire que toutes les variables ont une racine unitaire et sont intégrées d'ordre 1. Si cette relation est vraie dans la réalité, elle implique donc une cointégration entre primes en logarithmes, taux d'intérêt, indice boursier en logarithmes, sinistres en logarithmes, dépenses en logarithmes, capital en logarithmes, longueur et inflation. C'est ce qu'il faut vérifier sur les données observées de la réalité, en travaillant avec TSP. Le fichier Excel assur.xls téléchargeable sur le site www.pearsoneducation.fr contient les séries observées suivantes :

Primesem : les primes nominales totales ;
Obl : le taux d'intérêt nominal sur les obligations ;
Cac : l'indice boursier nominal ;
Sini : les sinistres nominaux totaux ;
Dep : les dépenses nominales de gestion ;

Exercices

Cp : le capital nominal des sociétés d'assurance ;
K : la longueur moyenne des branches ;
Inf : le taux d'inflation.
Ce fichier est constitué de manière à pouvoir être lu par TSP (voir chapitre 2).

Solution Pour réaliser le test de cointégration, il faut donner au logiciel TSP les instructions suivantes :

```
freq a;
smpl 1963 1999;
read(file='assur.xls');
lsin=log(sini);
lprimesem=log(primesem);
ldep=log(dep);
lcp=log(cp);
lcac=log(cac);
?
? tests engle granger de cointégration 65-99
?
smpl 1965 1999;
regopt(pvprint, lmlags=2, qlags=2) all;
regopt(noprint) @vcov @vcor;
noplots;
olsq lprimesem c obl lcac lsin ldep lcp k inf;
res=@res;
smpl 1966 1999;
dres=res-res(-1);
smpl 1967 1999;
olsq dres res(-1);
cdf(dickeyf, nvar=6, constant,notrend) @t(1);
olsq dres res(-1) dres(-1);
cdf(dickeyf,nvar=6,constant,notrend) @t(1);
```

Les résultats sont les suivants :

```
                              Equation 1
                             ============

               Method of estimation = Ordinary Least Squares

 Dependent variable: LPRIMESE
 Current sample: 1965 to 1999
 Number of observations: 35

  Mean of dependent variable = 11.3867
Std.  dev. of dependent var. = .995474
    Sum of squared residuals = .035663
       Variance of residuals = .132083E-02
    Std. error of regression = .036343
                   R-squared = .998942
```

```
          Adjusted R-squared = .998667
     Durbin-Watson statistic = .757481
  Breusch/Godfrey LM: AR/MA1 = 18.0745 [.000]
  Breusch/Godfrey LM: AR/MA2 = 22.6673 [.000]
      Ljung-Box Q-statistic1 = 12.6601 [.000]
      Ljung-Box Q-statistic2 = 12.6663 [.002]
   Wald nonlin. AR1 vs. lags = 12.3197 [.091]
                   ARCH test = 2.26463 [.132]
                  CuSum test = .376031 [.994]
                CuSumSq test = .466986 [.001]
                   Chow test = 8.25112 [.000]
      LR het. test (w/ Chow) = 42.5555 [.000]
  Jarque-Bera normality test = .537588 [.764]
   F-statistic (zero slopes) = 3640.26 [.000]
    Akaike Information Crit. = -3.59398
  Schwarz Bayes. Info. Crit. = -6.07635
  Log of likelihood function = 70.8947

           Estimated    Standard
Variable  Coefficient     Error       t-statistic   P-value
C         1.30534       .254761       5.12379       [.000]
OBL       .366363E-02   .662020E-02   .553402       [.585]
LCAC      -.031730      .050854       -.623944      [.538]
LSIN      -.777358E-02  .085036       -.091416      [.928]
LDEP      .937596       .109216       8.58475       [.000]
LCP       .026854       .068117       .394238       [.697]
K         .280884       .130808       2.14729       [.041]
INF      -.343361E-02   .465784E-02   -.737168      [.467]

Current sample:  1966 to 1999

Current sample:  1967 to 1999

                              Equation 2
                             ============

               Method of estimation = Ordinary Least Squares

Dependent variable:  DRES
Current sample:  1967 to 1999
Number of observations:  33

 Mean of dependent variable = .122208E-02
Std. dev. of dependent var. = .028996
   Sum of squared residuals = .020384
      Variance of residuals = .637005E-03
   Std. error of regression = .025239
                  R-squared = .247712
         Adjusted R-squared = .247712
    Durbin-Watson statistic = 1.17865
 Breusch/Godfrey LM: AR/MA1 = 13.1137 [.000]
 Breusch/Godfrey LM: AR/MA2 = 10.1149 [.006]
     Ljung-Box Q-statistic1 = 4.37589 [.036]
     Ljung-Box Q-statistic2 = 4.62909 [.099]
                  ARCH test = 1.40935 [.235]
```

```
                 CuSum test = .599497 [.417]
               CuSumSq test = .223404 [.203]
                  Chow test = .172019 [.681]
     LR het. test (w/ Chow) = -.817481 [1.00]
            White het. test = 0. [1.00]
 Jarque-Bera normality test = .095288 [.953]
   Akaike Information Crit. = -4.49102
Schwarz Bayes.  Info. Crit. = -7.28355
 Log of likelihood function = 75.1019

          Estimated     Standard
Variable  Coefficient     Error       t-statistic   P-value
RES(-1)   -.449767       .140060      -3.21124      [.003]
DICKEY-FULLER(C,ASY.,6) Test Statistic: -3.211236, Lower tail area: .63399

                               Equation 3
                              ============

               Method of estimation = Ordinary Least Squares

Dependent variable:  DRES
Current sample:  1967 to 1999
Number of observations:  33

 Mean of dependent variable = .122208E-02
Std. dev. of dependent var. = .028996
   Sum of squared residuals = .013656
      Variance of residuals = .440512E-03
   Std. error of regression = .020988
                  R-squared = .500584
         Adjusted R-squared = .484474
    Durbin-Watson statistic = 1.94212
                 Durbin's h = -.550791 [.582]
     Durbin's h alternative = -.745214 [.456]
 Breusch/Godfrey LM: AR/MA1 = .555344 [.456]
 Breusch/Godfrey LM: AR/MA2 = 1.12597 [.570]
     Ljung-Box Q-statistic1 = .072370 [.788]
     Ljung-Box Q-statistic2 = 1.36287 [.506]
                  ARCH test = .144538 [.704]
                 CuSum test = .697415 [.253]
               CuSumSq test = .179523 [.425]
                  Chow test = .421788 [.660]
     LR het. test (w/ Chow) = -1.22798 [1.00]
            White het. test = 0. [1.00]
 Jarque-Bera normality test = .945372 [.623]
  F-statistic (zero slopes) = 30.0736 [.000]
   Akaike Information Crit. = -4.83100
Schwarz Bayes.  Info. Crit. = -7.57818
 Log of likelihood function = 81.7115

            Estimated    Standard
Variable   Coefficient     Error     t-statistic   P-value
RES(-1)    -.672522      .129671     -5.18638      [.000]
DRES(-1)   .584632       .149592     3.90816       [.000]
DICKEY-FULLER(C,ASY.,6) Test Statistic : -5.186382, Lower tail area: .01221
```



Les résultats présentés sous le titre « equation 1 » sont produits par l'instruction `olsq lprimesem c obl lcac lsin ldep lcp k inf;`. Les coefficients estimés fournissent donc une estimation du vecteur de cointégration (les β_i) entre `lprimesem` et `obl`, `lcac`, `lsin`, `ldep`, `lcp`, `k`, et `inf`, en cas de réelle cointégration entre ces variables. Pour le vérifier, il faut tester si le résidu de la régression linéaire est stationnaire, c'est-à-dire si l'on rejette l'hypothèse que ce résidu a une racine de valeur 1.

C'est la raison pour laquelle on définit une variable `res` égale au résidu de la régression, au moyen de l'instruction `res=@res`. En effet, après avoir calculé une régression par MCO, le logiciel TSP garde automatiquement en mémoire la série des résidus sous le nom de `@res`. Ici, on renomme cette série en l'appelant `res`. Ensuite, on calcule sa variation, appelée `dres`, au moyen de l'instruction `dres=res-res(-1)`.

Les résultats présentés sous le titre « equation 2 » sont produits par l'instruction `olsq dres res(-1)`. Elle calcule par MCO la régression de base sur laquelle se fonde le test *DF*. Il correspond à la *t*-stat du coefficient de la variable `res(-1)` dans cette régression, soit à la valeur de `@t(1)`. Après avoir calculé une régression par MCO, le logiciel TSP garde automatiquement en mémoire le vecteur des *t*-stats sous le nom de `@t`. Le premier élément de ce vecteur est `@t(1)`, le deuxième `@t(2)`, etc. La *t*-stat du coefficient de la variable `res(-1)` est le premier élément de `@res`, soit `@t(1)`. Il faut comparer cette valeur à une table de Dickey et Fuller pour déterminer si l'on rejette ou non l'hypothèse de racine unitaire pour la variable `res`. L'instruction `cdf(dickeyf,nvar=6,constant,notrend) @t(1)` se charge de l'opération. L'argument `dickeyf` compare `@t(1)` à une table adéquate de Dickey et Fuller. L'argument `nvar=6` utilise la table valable pour une relation entre six variables (on ne dispose pas de table du test pour huit variables, six étant le maximum existant). Les arguments `constant` et `notrend` indiquent la présence d'une constante et l'absence de tendance déterministe dans l'équation. Le résultat de `cdf(dickeyf,nvar=6,constant,notrend) @t(1);` est la ligne : `db DICKEY-FULLER(C,ASY.,6) Test Statistic: -3.211236, Lower tail area: .63399`

Le test *DF* est −3.211236. Puisque la probabilité critique est .63399, soit une valeur supérieure à 0,05, et même à 0,10, on ne rejette pas l'hypothèse de racine unitaire pour la variable `res` aux seuils de signification de 5 % et de 10 %, d'après ce test. Toutefois, on ne sait pas encore si le test *DF* est fiable ou s'il faut plutôt s'en remettre à un test *ADF*.

Les résultats présentés sous le titre « equation 3 » sont produits par l'instruction `olsq dres res(-1) dres(-1)`. Elle calcule par MCO la régression de base sur laquelle se fonde le test ADF(1). Il correspond à la t-stat du coefficient de la variable `res(-1)`, soit à `@t(1)`. Il faut de nouveau comparer cette valeur à une table de Dickey et Fuller pour déterminer si l'on rejette ou non l'hypothèse de racine unitaire pour la variable `res`. L'instruction `cdf(dickeyf,nvar=6,constant,notrend) @t(1)` se charge de l'opération et produit comme résultat la ligne :

```
DICKEY-FULLER(C,ASY.,6) Test Statistic : -5.186382, Lower tail area : .01221
```

Le test *ADF*(1) est −5.186382. Puisque la probabilité critique est .01221, soit une valeur inférieure à 0,05, on rejette l'hypothèse de racine unitaire au seuil de signification de 5 %, d'après ce test.

Il faut se fier au test *ADF*(1) plutôt qu'au test *DF*. En effet, le logarithme de la fonction de vraisemblance est 81.7115, contre 75.1019 précédemment. Le test du rapport de vraisemblance, $2\big(\ln(L_1) - \ln(L_0)\big) = 2(81.7115 - 75.1019)$, est supérieur à la valeur critique d'une Chi-2 à 1 degré de liberté : on rejette donc la spécification sans la variable `dres(-1)` au profit de celle qui l'inclut. La probabilité critique .000 de la *t*-stat du

Exercices

coefficient de `dres(-1)` est d'ailleurs inférieure à 0,05 et implique donc que le coefficient de `dres(-1)` n'est pas 0. Il faut ajouter la variable `dres(-1)` pour que le terme d'erreur de la régression auxiliaire sur le résidu soit un bruit blanc. En effet, dans l'équation du test DF qui n'inclut pas `dres(-1)`, les probabilités critiques des tests d'autocorrélation « Breusch/Godfrey LM » et « Ljung-Box Q-statistic » sont toutes inférieures à 0,05 : on rejette donc l'hypothèse d'absence d'autocorrélation. Par contre, dans l'équation du test *ADF*(1) qui inclut la variable `dres(-1)`, ces mêmes probabilités critiques sont toutes supérieures à 0,05 : on ne rejette pas l'hypothèse en question. Un autre argument en faveur du test *ADF*(1) est que le critère d'Akaike de l'équation du test *ADF*(1) qui inclut la variable `dres(-1)` (−4.83100) prend une valeur supérieure à celle du même critère (−4.49102) de l'équation du test *DF* qui n'inclut pas `dres(-1)`.

Puisqu'on doit se fier au test *ADF*(1), et que celui-ci indique que la variable `res` n'a pas de racine unitaire, donc qu'elle est stationnaire, on conclut à une cointégration entre les variables.

Le logiciel TSP est capable de réaliser les tests précédents. Les instructions à lui faire exécuter sont les suivantes :

```
freq a;
smpl 1963 1999;
read(file='assur.xls');
lsin=log(sini);
lprimesem=log(primesem);
ldep=log(dep);
lcp=log(cp);
lcac=log(cac);
?
? tests engle granger de cointégration 65-99
?
smpl 1965 1999;
coint(eg,nojoh,maxlag=1,constant,notrend)
lprimesem obl lcac lsin ldep lcp k inf;
```

Les résultats sont les suivants :

```
Current sample:  1963 to 1999
Current sample: 1965 to 1999

Engle-Granger (tau) cointegration tests
=======================================

Variables: LPRIMESEM OBL LCAC LSIN LDEP LCP K INF

                         Dependent variable: LPRIMESEM

Num lags              0              1            Opt:1
alpha           0.55023        0.32748          0.32748
TestStat       -3.21124       -5.18638         -5.18638
P-value           .              .                .
Const           1.30534        1.30534          1.30534
```




```
t(Const)        5.12379         5.12379          5.12379
Num obs        33.00000        33.00000         33.00000
LogLike        75.10186        81.71155         81.71155
AIC            -4.49102        -4.83100         -4.83100
Var res      0.00063700      0.00044051       0.00044051

                                 Cointegrating vector

                       1
LPRIMESEM            1.00000
OBL               -0.0036636
LCAC                0.031730
LSIN               0.0077736
LDEP                -0.93760
LCP                -0.026854
K                   -0.28088
INF                0.0034336

              Summary of Engle-Granger (tau) cointegration tests

             TestStat       P-value      Num.lags
1            -5.18638          .           1.00000
```

La colonne 0 reprend les résultats du test *DF*, la colonne 1 ceux du test *ADF*(1). Le test *DF* est donc la valeur de TestStat de la colonne 0 : −3.21124, tandis que le test *ADF*(1) est la valeur de TestStat de la colonne 1 : −5.18638. Ces tests sont exactement égaux à ceux obtenus par la procédure manuelle décrite précédemment, dans les équations 2 et 3. Const et t(Const) correspondent respectivement à la constante estimée dans la relation de cointégration et à la t-stat de cette constante ; les valeurs de l'une comme de l'autre sont celles de l'équation 1 dans la procédure manuelle. Les valeurs de LogLike sont celles de Log of likelihood function dans les équations 2 et 3. Les valeurs de AIC sont celles du critère d'Akaike dans ces mêmes équations. Les probabilités critiques ne sont pas imprimées par TSP (il n'existe pas de table pour une relation à huit variables). Toutefois, en utilisant celle à six variables, on rejette l'hypothèse de racine unitaire pour le résidu de la relation estimée (cointégration entre les variables).

Exercice 2 Assurance et modèle à correction d'erreur

Énoncé

Sur la base de l'estimation du vecteur cointégrant de l'exercice précédent, spécifiez et estimez un modèle à correction d'assurance pour le logarithme des primes. Utilisez les mêmes données que précédemment. Travaillez dans TSP.

Solution

Pour estimer un modèle à correction d'erreur pour les primes totales en logarithmes, on envoie à TSP les instructions suivantes :

```
freq a;
smpl 1963 1999;
```

```
read(file='assur.xls');
?
? calcul des transformations logarithmiques
?
lsin=log(sini);
lprimesem=log(primesem);
ldep=log(dep);
lcp=log(cp);
lcac=log(cac);
?
? calcul des différences
?
smpl 1964 1999;
dlprimes=lprimesem-lprimesem(-1);
dldep=ldep-ldep(-1);
dobl=obl-obl(-1);
dlcac=lcac-lcac(-1);
dlsin=lsin-lsin(-);
dlcp=lcp-lcp(-1);
dinf=inf-inf(-1);
dk=k-k(-1);
?
? estimation de la relation de cointégration
?
smpl 1965 1999;
regopt(pvprint,lmlags=2,qlags=2) all;
regopt(noprint) @vcov @vcor;
noplots;
olsq lprimesem c obl lcac lsin ldep lcp k inf;
?
? sauvegarde des résidus qui sont l'écart à l'équilibre
?
res=@res;
?
? estimation du modèle à correction d'erreur
?
smpl 1966 1999;
olsq dlprimes c res(-1)
dlprimes(-1) dobl(-1) dlcac(-1) dlsin(-1) dldep(-1) dlcp(-1) dk(-1) dinf(-1)
dlprimes(-2) dobl(-2) dlcac(-2) dlsin(-2) dldep(-2) dlcp(-2) dk(-2) dinf(-2);
```

Les commentaires insérés dans le programme facilitent la compréhension des opérations successives. Ici on a choisi arbitrairement deux retards pour toutes les variations des variables. Les résultats sont les suivants :

```
                              Equation 2
                             ============

                Method of estimation = Ordinary Least Squares

Dependent variable:  DLPRIMES
Current sample:  1966 to 1999
Number of observations:  34

  Mean of dependent variable = .089959
Std.  dev. of dependent var. = .050749
```

```
    Sum of squared residuals = .011828
       Variance of residuals = .739252E-03
    Std. error of regression = .027189
                   R-squared = .860829
          Adjusted R-squared = .712959
     Durbin-Watson statistic = 2.03359
      Durbin's h alternative = .118536 [.906]
  Breusch/Godfrey LM: AR/MA1 = .014051 [.906]
  Breusch/Godfrey LM: AR/MA2 = 4.13464 [.127]
      Ljung-Box Q-statistic1 = .118590 [.731]
      Ljung-Box Q-statistic2 = .305927 [.858]
                   ARCH test = .047164 [.828]
                  CuSum test = .532840 [.561]
                CuSumSq test = .155000 [.807]
  Jarque-Bera normality test = .043710 [.978]
   F-statistic (zero slopes) = 5.82154 [.000]
    Akaike Information Crit. = -4.06694
  Schwarz Bayes. Info. Crit. = -6.09675
  Log of likelihood function = 87.1380
```

Variable	Estimated Coefficient	Standard Error	t-statistic	P-value
C	-.576536E-02	.017461	-.330193	[.746]
RES(-1)	-.471463	.250898	-1.87910	[.079]
DLPRIMES(-1)	.400682	.240846	1.66365	[.116]
DOBL(-1)	.421109E-02	.928274E-02	.453647	[.656]
DLCAC(-1)	-.548736E-02	.062097	-.088368	[.931]
DLSIN(-1)	.016499	.111949	.147383	[.885]
DLDEP(-1)	.086245	.329789	.261516	[.797]
DLCP(-1)	.015883	.097130	.163524	[.872]
DK(-1)	.162739	.148487	1.09599	[.289]
DINF(-1)	.583384E-02	.584087E-02	.998797	[.333]
DLPRIMES(-2)	.420576	.236128	1.78113	[.094]
DOBL(-2)	.141081E-02	.010293	.137067	[.893]
DLCAC(-2)	.011794	.049593	.237812	[.815]
DLSIN(-2)	.073032	.106076	.688490	[.501]
DLDEP(-2)	-.038658	.280898	-.137621	[.892]
DLCP(-2)	-.265483E-02	.102018	-.026023	[.980]
DK(-2)	-.020630	.200676	-.102803	[.919]
DINF(-2)	-.279242E-02	.573218E-02	-.487149	[.633]

Le coefficient de `res(-1)` exprime le degré de variation du logarithme des primes, à la période courante, suite à une augmentation unitaire, à la période précédente, de l'écart entre le logarithme des primes et sa valeur d'équilibre. Une telle augmentation implique que les primes étaient à l'époque trop élevées par rapport à leur valeur d'équilibre. Pour corriger cet écart, il faut diminuer les primes. En toute logique, le coefficient estimé de `res(-1)` devrait être négatif.

Puisque le nombre de variables explicatives est conséquent, la multicolinéarité est certainement importante, ce qui explique que les coefficients ne peuvent pas être estimés

avec précision, que leurs écarts types sont par conséquent élevés, et par là même que leurs *t*-stats sont très faibles. Il faut simplifier ce modèle en supprimant les variables explicatives inutiles. On peut vraisemblablement supprimer toutes les variations retardées de deux périodes, sauf celle des primes, et presque toutes les variations retardées de une période. Pour estimer de nouveau le modèle avec ces restrictions, on envoie à TSP les instructions suivantes :

```
freq a;
smpl 1963 1999;
read(file='assur.xls');
?
? calcul des transformations logarithmiques
?
lsin=log(sini);
lprimesem=log(primesem);
ldep=log(dep);
lcp=log(cp);
lcac=log(cac);
?
? calcul des différences
?
smpl 1964 1999;
dlprimes=lprimesem-lprimesem(-1);
dldep=ldep-ldep(-1);
dobl=obl-obl(-1);
dlcac=lcac-lcac(-1);
dlsin=lsin-lsin(-1);
dlcp=lcp-lcp(-1);
dinf=inf-inf(-1);
dk=k-k(-1);
?
? estimation de la relation de cointégration
?
smpl 1965 1999;
regopt(pvprint,lmlags=2,qlags=2) all;
regopt(noprint) @vcov @vcor;
noplots;
olsq lprimesem c obl lcac lsin ldep lcp k inf;
?
? sauvegarde des résidus qui sont l'écart à l'équilibre
?
res=@res;
?
? estimation du modèle à correction d'erreur
?
smpl 1966 1999;
olsq dlprimes c res(-1) dlprimes(-1) dinf(-1) dlprimes(-2);
```



Les résultats sont les suivants :

```
                              Equation 2
                             ============

                  Method of estimation = Ordinary Least Squares

Dependent variable:  DLPRIMES
Current sample:  1966 to 1999
Number of observations:  34

 Mean of dependent variable = .089959
Std. dev. of dependent var. = .050749
   Sum of squared residuals = .014615
      Variance of residuals = .503952E-03
   Std. error of regression = .022449
                  R-squared = .828041
         Adjusted R-squared = .804323
    Durbin-Watson statistic = 2.05801
     Durbin's h alternative = -.277695 [.781]
 Breusch/Godfrey LM: AR/MA1 = .077115 [.781]
 Breusch/Godfrey LM: AR/MA2 = 2.02168 [.364]
     Ljung-Box Q-statistic1 = .437181 [.508]
     Ljung-Box Q-statistic2 = 1.05639 [.590]
                  ARCH test = 1.40040 [.237]
                 CuSum test = .511158 [.614]
               CuSumSq test = .201041 [.333]
                  Chow test = 2.52281 [.057]
     LR het. test (w/ Chow) = 13.2146 [.000]
           White het.  test = 18.4245 [.188]
 Jarque-Bera normality test = .423232 [.809]
  F-statistic (zero slopes) = 34.9112 [.000]
  Akaike Information Crit.  = -4.62010
Schwarz Bayes.  Info. Crit. = -7.23351
 Log of likelihood function = 83.5417

               Estimated     Standard
Variable      Coefficient      Error        t-statistic   P-value
C             .333048E-02   .931957E-02   .357364       [.723]
RES(-1)       -.600287      .123689       -4.85320      [.000]
DLPRIMES(-1)  .465483       .128287       3.62846       [.001]
DINF(-1)      .729985E-02   .241583E-02   3.02168       [.005]
DLPRIMES(-2)  .449595       .134504       3.34261       [.002]
```

Exercice 3 Assurance et tests de Johansen

Énoncé

Spécifiez et estimez un modèle explicatif de l'évolution temporelle du total annuel des primes perçues par les sociétés françaises d'assurance en dommages. Plus précisément, expliquez l'évolution des primes réelles. Le logarithme de celles-ci est noté $LPRIMESR_t$ et est défini par $LPRIMESR_t = \ln(PRIMESEM_t/PRIX_t)$, où la variable PRIX_t est l'indice

des prix à la consommation. Toutes les données nécessaires se trouvent dans le fichier ASSUR.xls. Les variables explicatives potentielles sont :

– le logarithme du capital réel des sociétés d'assurance : $LCPR_t = \ln(CP_t/PRIX_t)$;
– le logarithme des dépenses réelles : $LDEPR_t = \ln(DEP_t/PRIX_t)$;
– le logarithme du PIB réel : $LPIBR_t = \ln(PIBLEVEL_t)$;
– le taux d'intérêt réel : $OBLR_t = OBL_t - INF_t$;
– le taux de croissance de l'indice boursier réel : $DLCACR_t = \ln(CAC_t/PRIX_t) - \ln(CAC_{t-1}/PRIX_{t-1})$;
– le logarithme des sinistres réels : $LSINR_t = \ln(SINI_t/PRIX_t)$.

Il est facile de vérifier par des tests de racine unitaire que toutes ces variables sont intégrées d'ordre 1. Elles ne peuvent donc entretenir qu'une relation de cointégration. Testez cette éventuelle cointégration par la méthode de Johansen. Utilisez un $VAR(2)$.

Solution

Toutes les variables forment le processus stochastique multivarié $\{X_t\}$, défini de la manière suivante :

$$X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \\ X_{4t} \\ X_{5t} \\ X_{6t} \\ X_{7t} \end{pmatrix} = \begin{pmatrix} lprimesr_t \\ lcpr_t \\ ldepr_t \\ lpibr_t \\ oblr_t \\ \Delta lcacr_t \\ lsinr_t \end{pmatrix}$$

Le processus $\{X_t\}$ est censé être déterminé par un système dynamique stochastique susceptible d'être représenté par un modèle $VAR(2)$:

$$X_t = \mu + \sum_{l=1}^{2} \Phi_l X_{t-l} + u_t$$

c'est-à-dire que :

$$\Phi(L)X_t = \mu + u_t$$

où $\Phi(L)$ est un polynôme matriciel en l'opérateur de retard L, défini par :

$$\Phi(L) = \left(I_7 - \sum_{l=1}^{2} \Phi_l L^l \right)$$

Chaque matrice Φ_l est d'ordre 7×7. Le choix d'un VAR à deux retards est justifié par le fait qu'une dynamique à deux retards est bien adaptée à des séries annuelles : elle permet la présence de cycles dans les variables. L'existence de racines unitaires, établie par les tests précédents, implique que l'équation caractéristique $(0 = |\Phi(L)|)$ du système VAR admet des racines unitaires et donc que $0 = |\Phi(1)|$. Ainsi, la matrice $\Phi(1) = \left(I_7 - \sum_{l=1}^{2} \Phi_l\right)$ est singulière.

Par ailleurs, on montre aisément que $\Phi(L)$ se décompose de la manière suivante :

$$\Phi(L) = \Phi(1)L + (1 - L)(I_n - A_1 L)$$

où $A_1 = -\Phi_2$.

Par conséquent, le système $VAR(2)$, dont la forme originale est $\Phi(L)X_t = \mu + u_t$, peut encore s'écrire de la manière suivante :

$$\Delta X_t = \mu - \Phi(1)X_{t-1} + A_1\Delta X_{t-1} + u_t$$

Les variables de X sont cointégrées, avec r vecteurs de cointégration linéairement indépendants si le rang de $\Phi(1)$ est égal à r ($< n$) et s'il existe des matrices α et β, de type $7 \times r$ et de rang r, où $0 < r \leqslant n$, telles que $\Phi(1) = -\alpha\beta'$. On peut alors représenter le système $VAR(2)$ sous la forme d'un mécanisme à correction d'erreur, de la manière suivante :

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + A_1\Delta X_{t-1} + u_t$$

La matrice de cointégration β est de type $7 \times r$ et telle que chaque colonne est un vecteur de cointégration distinct des autres. Le vecteur $\beta' X_{t-1}$, qui a r éléments, est le vecteur des r expressions cointégrées qui mesurent chacune un écart par rapport à l'une des relations d'équilibre (à une expression correspond une relation, et inversement), à la période $t-1$. Si l'on explicite les matrices, la forme à correction d'erreur est :

$$\begin{pmatrix} \Delta primesr_t \\ \Delta cpr_t \\ \vdots \\ \Delta sinir_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_7 \end{pmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1r} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2r} \\ \cdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nr} \end{bmatrix}$$

$$\times \begin{pmatrix} \beta_{11} primesr_{t-1} + \beta_{21} cpr_{t-1} + \cdots + \beta_{71} sinir_{t-1} \\ \beta_{12} primesr_{t-1} + \beta_{22} cpr_{t-1} + \cdots + \beta_{72} sinir_{t-1} \\ \vdots \\ \beta_{1r} primesr_{t-1} + \beta_{2r} cpr_{t-1} + \cdots + \beta_{7r} sinir_{t-1} \end{pmatrix}$$

$$+ \begin{pmatrix} a_{11,1} & a_{12,1} & \cdots & a_{17,1} \\ a_{21,1} & a_{22,1} & \cdots & a_{27,1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{71,1} & a_{72,1} & \cdots & a_{77,1} \end{pmatrix} \begin{pmatrix} \Delta primesr_{t-1} \\ \Delta cpr_{t-1} \\ \vdots \\ \Delta sinir_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{7t} \end{pmatrix}$$

En cas de cointégration, le système $VAR(2)$ cointégré à correction d'erreur est donc le système d'équations suivantes pour les sept variables :

$$\begin{aligned} \Delta lprimesr_t = \mu_1 &+ \alpha_{11}\left(\beta_{11} lprimesr_{t-1} + \beta_{21} lcpr_{t-1} + \cdots + \beta_{71} lsinr_{t-1}\right) \\ &+ \alpha_{12}\left(\beta_{12} lprimesr_{t-1} + \beta_{22} lcpr_{t-1} + \cdots + \beta_{72} lsinr_{t-1}\right) + \cdots \\ &+ \alpha_{1r}\left(\beta_{1r} lprimesr_{t-1} + \beta_{2r} lcpr_{t-1} + \cdots + \beta_{7r} lsinr_{t-1}\right) \\ &+ a_{11,1}\Delta lprimesr_{t-1} + a_{12,1}\Delta lcpr_{t-1} + \cdots + a_{17,1}\Delta lsinr_{t-1} + u_{1t} \\ \Delta cpr_t = \mu_2 &+ \alpha_{21}\left(\beta_{11} primesr_{t-1} + \beta_{21} cpr_{t-1} + \cdots + \beta_{71} sinir_{t-1}\right) \\ &+ \alpha_{22}\left(\beta_{12} primesr_{t-1} + \beta_{22} cpr_{t-1} + \cdots + \beta_{72} sinir_{t-1}\right) + \cdots \\ &+ \alpha_{2r}\left(\beta_{1r} primesr_{t-1} + \beta_{2r} cpr_{t-1} + \cdots + \beta_{7r} sinir_{t-1}\right) \\ &\times a_{21,1}\Delta primesr_{t-1} + a_{22,1}\Delta cpr_{t-1} + \cdots + a_{27,1}\Delta sinir_{t-1} + u_{2t} \end{aligned}$$

$$\vdots$$

$$\begin{aligned}\Delta sinir_t = \mu_7 + \alpha_{71}\left(\beta_{11} primesr_{t-1} + \beta_{21} cpr_{t-1} + \cdots + \beta_{71} sinir_{t-1}\right)\\ + \alpha_{72}\left(\beta_{12} primesr_{t-1} + \beta_{22} cpr_{t-1} + \cdots + \beta_{72} sinir_{t-1}\right) + \cdots\\ + \alpha_{7r}\left(\beta_{1r} primesr_{t-1} + \beta_{2r} cpr_{t-1} + \cdots + \beta_{7r} sinir_{t-1}\right)\\ \times a_{71,1}\Delta primesr_{t-1} + a_{72,1}\Delta cpr_{t-1} + \cdots + a_{77,1}\Delta sinir_{t-1} + u_{7t}\end{aligned}$$

Chaque coefficient α_{ij} de la matrice α représente le taux de réaction de la i^e variable à l'écart précédent par rapport à la j^e relation d'équilibre. Il s'agit donc du taux de correction de la variable i, suite à l'observation de l'écart (ou déséquilibre) précédent, par rapport à la j^e relation d'équilibre. Chaque coefficient $\alpha_{il,1}$ de la matrice A_1 représente le taux de réaction de la i^e variable à la variation passée, à la période précédente, de la l^e variable.

Il faut vérifier la cointégration, c'est-à-dire que r est supérieur à 0, et identifier quelle est la valeur précise de r.

Réalisation avec les logiciels

Pour tester [1] la cointégration entre *LPRIMESR*, *LCPR*, *LDEPR*, *LPIBR*, *OBLR*, *DLCACR* et *LSINR* par la méthode de Johansen, avec un *VAR*(2) sans tendance déterministe ajoutée et sans restriction sur μ, on utilise les instructions TSP suivantes :

```
SMPL 1963 1999;
FREQ A;
READ(FILE='C :\ASSUR.XLS');
LPRIMESR=LOG(PRIMESEM/PRIX);
LCPR=LOG(CP/PRIX);
LDEPR=LOG(DEP/PRIX);
LSINR=LOG(SINI/PRIX);
LPIBR=LOG(PIBLEVEL);
OBLR=OBL-INF;
SMPL 1964 1999;
DLCACR=LOG(CAC/PRIX)-LOG(CAC(-1)/PRIX(-1));
COINT(NOUNIT,JOH,NoeG, MINLAG=1, MAXLAG=1, CONST, NOTREND) LPRIMESR LCPR LDEPR LPIBR
OBLR DLCACR LSINR;
```

L'option `MAXLAG=1` commande l'utilisation d'un *VAR*(2), puisque `MAXLAG` est le nombre de retards des différences des variables dans le *VAR* à correction d'erreur. Les résultats sont les suivants :

```
Johansen (trace) cointegration tests
====================================

Variables: LPRIMESR LCPR LDEPR LPIBR OBLR DLCACR LSINR

Num lags          1
Eigval1     0.86788
Eigval2     0.72159
Eigval3     0.67427
```

1. Le rang de cointégration est déterminé par un test sur la nullité des $(7 - r)$ plus petites valeurs propres, ce qui correspond à un test de rapport de vraisemblance.



```
Eigval4      0.41006
Eigval5      0.35690
Eigval6      0.20504
Eigval7     0.021400
H0:r=0     112.89414
P-valAsy     0.20319
H0:r<=1     72.41334
P-valAsy     0.62619
H0:r<=2     46.83983
P-valAsy     0.77153
H0:r<=3     24.40583
P-valAsy     0.89665
H0:r<=4     13.85115
P-valAsy     0.83798
H0:r<=5      5.02197
P-valAsy     0.82994
H0:r<=6      0.43265
P-valAsy     0.76107
Num obs     34.00000
LogLike    414.54897
AIC       -309.54897

                   Cointegrating vectors for lag 1

                   1            2             3             4
LPRIMESR     1.00000      1.00000       1.00000       1.00000
LCPR         0.22032    -0.015788      -0.11433  -0.000081383
LDEPR       -0.54291     -3.14043      -0.78067       0.39600
LPIBR        0.45399      2.92425       0.47101      -2.07179
OBLR       -0.020367     0.024538   -0.00022052     0.0091099
DLCACR     -0.081621     -0.35504      -0.75793      0.096830
LSINR       -0.83078     -0.30152      0.011665       0.13164

                   5            6             7
LPRIMESR     1.00000      1.00000       1.00000
LCPR        -0.55720     -0.48067     -0.087056
LDEPR        1.25083     -6.44803      -4.46339
LPIBR        1.69603      6.43380       3.95079
OBLR        0.055339    -0.061996     0.0039189
DLCACR       0.16266      0.30600      0.018163
LSINR       -1.88432      1.22627       0.52168
```

Les valeurs du test de la trace sont inférieures à celles données par les formules originales de Johansen, en raison de la correction pour de petits échantillons, utilisée par TSP. Le test de la plus grande valeur propre n'est pas disponible dans la procédure automatique de TSP, mais peut être programmé aisément.

Les `Eigval` sont les valeurs propres classées par ordre décroissant. Viennent ensuite les valeurs du test de Johansen pour les hypothèses successives sur la valeur de *r*, avec leurs probabilités critiques. Celles-ci sont interpolées à partir des valeurs critiques des tables

d'Osterwald-Lenum [OST 1992]. Les valeurs du test sont très élevées pour les petits r ; à partir de l'hypothèse $r \leqslant 3$, elles deviennent suffisamment petites. Si l'on décide de ne pas rejeter l'hypothèse que $r = 3$, les vecteurs cointégrants correspondants sont les trois premières colonnes du tableau des « cointegrating vectors ».

Avec d'autres logiciels, on obtient les résultats suivants pour les tests originaux de Johansen, sur les mêmes données, en utilisant de nouveau un *VAR*(2), sans tendance déterministe ajoutée et sans restriction sur la constante.

Tableau 6.1

Test fondé sur la valeur propre maximale

H0	H1	Test	Valeur Critique à 95 %	Valeur Critique à 90 %
r = 0	r = 1	68.7972	45.6300	42.7000
r<= 1	r = 2	43.5940	39.8300	36.8400
r<= 2	r = 3	38.1290	33.6400	31.0200
r<= 3	r = 4	17.8790	27.4200	24.9900
r<= 4	r = 5	15.0292	21.1200	19.0200
r<= 5	r = 6	7.7850	14.8800	12.9800
r<= 6	r = 7	0.7393	8.0700	6.5000

Tableau 6.2

Test fondé sur la trace

H0	H1	Test	Valeur Critique à 95 %	Valeur Critique à 90 %
r = 0	r>= 1	191.9527	124.6200	119.6800
r<= 1	r>= 2	123.1555	95.8700	91.4000
r<= 2	r>= 3	79.5615	70.4900	66.2300
r<= 3	r>= 4	41.4325	48.8800	45.7000
r<= 4	r>= 5	23.5535	31.5400	28.7800
r<= 5	r>= 6	8.5243	17.8600	15.7500
r<= 6	r = 7	0.7393	8.0700	6.5000

On utilise ici les valeurs critiques définies par la méthode de Pesaran, Shin et Smith [PES 1996] plutôt que celles définies par Johansen ou par Osterwald-Lenum. À partir de l'hypothèse $r \leqslant 3$, les valeurs du test sont inférieures aux valeurs critiques. Ces résultats suggèrent l'existence de trois relations de cointégration entre les variables étudiées. Sous cette restriction ($r = 3$), les vecteurs de cointégration estimés sont les suivants :

Les valeurs entre parenthèses sont les vecteurs estimés normalisés lorsqu'on impose la valeur -1 au premier coefficient. Mais il ne s'agit que d'une normalisation provisoire pour faciliter une première lecture des résultats. On voit que ces vecteurs normalisés sont quasiment identiques aux trois premiers vecteurs trouvés par TSP, aux colonnes 1 à 3.

La méthode de Johansen permet donc d'obtenir trois vecteurs de cointégration estimés. Toutefois, ils ne sont pas identifiés puisque toute combinaison linéaire de ces vecteurs estimés peut aussi être considérée comme un vecteur de cointégration. Pour les identifier, il faut imposer des restrictions d'identification dans le but de choisir des vecteurs $\tilde{\hat{\beta}}_1$, $\tilde{\hat{\beta}}_2$, $\tilde{\hat{\beta}}_3$ particuliers (qui sont les colonnes d'une matrice $\tilde{\hat{\beta}}$) parmi l'infinité de vecteurs qui sont des combinaisons linéaires des vecteurs estimés $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ (ceux-ci étant les colonnes de la matrice $\hat{\beta}$).

Les restrictions de normalisation choisies sont :

$$\tilde{\hat{\beta}}_{ii} = -1 \quad \forall i = 1, \ldots, 3$$



Tableau 6.3

Vecteurs de cointégration estimés sous l'hypothèse $r = 3$

	Vecteur 1	Vecteur 2	Vecteur 3
$primesr_t$	-3.1902	2.9196	-2.1187
	(-1.0000)	(-1.0000)	(-1.0000)
cpr_t	-.70301	-.047764	.24429
	(-.22037)	(.016360)	(.11530)
$depr_t$	1.7356	-9.1149	1.6146
	(.54404)	(3.1220)	(.76207)
$pibr_t$	-1.4528	8.4761	-.95198
	(-.45540)	(-2.9032)	(-.44932)
r_t	.065018	.070862	.8165E-3
	(.020381)	(-.024271)	(.3854E-3)
$\Delta cacr_t$	.26140	-1.0366	1.6153
	(.081938)	(.35504)	(.76241)
$sinir_t$	2.6504	-.87012	-.034227
	(.83081)	(.29803)	(-.016154)

Ces normalisations suggèrent implicitement que l'on souhaite interpréter le premier vecteur de cointégration comme une relation portant sur la détermination des primes, le deuxième vecteur de cointégration comme une relation portant sur la détermination du capital, et le troisième comme une relation portant sur la détermination des dépenses.

Des restrictions supplémentaires d'identification doivent également être imposées. Pour que les coefficients cointégrants soient tous exactement identifiés, il faut imposer deux restrictions sur les coefficients de chaque vecteur normalisé.

Ici, il semble raisonnable d'imposer que le PIB n'a pas d'influence directe sur les primes à long terme (il a un impact indirect en raison de son influence éventuelle sur les sinistres, qui affectent de manière essentielle l'évolution tendancielle des primes en question). Par ailleurs, à long terme, l'élasticité des primes réelles aux sinistres réels ne peut être qu'unitaire, pour respecter la cohérence théorique macroéconomique et financière.

On peut aussi raisonnablement supposer que l'évolution du capital des assureurs n'est pas liée directement au PIB réel, ni au taux d'intérêt réel à long terme.

Quant aux dépenses réelles des assureurs, elles n'ont pas de raison d'entretenir une relation d'équilibre spécifique avec le taux d'intérêt réel et le taux de rendement réel des actions.

On choisit donc des vecteurs $\tilde{\hat{\beta}}_1, \tilde{\hat{\beta}}_2, \tilde{\hat{\beta}}_3$ qui sont des combinaisons linéaires des vecteurs originaux $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$:

$$\tilde{\hat{\beta}}_1 = \hat{\beta}_1 W_1\,, \quad \tilde{\hat{\beta}}_2 = \hat{\beta}_2 W_2\,, \quad \tilde{\hat{\beta}}_3 = \hat{\beta}_3 W_3$$

où les poids $W_1 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix}$, $W_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}$ et $W_3 = \begin{pmatrix} a_3 \\ b_3 \\ c_3 \end{pmatrix}$ des combinaisons linéaires

doivent être sélectionnés sous les contraintes suivantes :

$$\begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} & \hat{\beta}_{13} \\ \hat{\beta}_{41} & \hat{\beta}_{42} & \hat{\beta}_{43} \\ \hat{\beta}_{71} & \hat{\beta}_{72} & \hat{\beta}_{73} \end{pmatrix} \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} \hat{\beta}_{21} & \hat{\beta}_{22} & \hat{\beta}_{23} \\ \hat{\beta}_{41} & \hat{\beta}_{42} & \hat{\beta}_{43} \\ \hat{\beta}_{51} & \hat{\beta}_{52} & \hat{\beta}_{53} \end{pmatrix} \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} \hat{\beta}_{31} & \hat{\beta}_{32} & \hat{\beta}_{33} \\ \hat{\beta}_{51} & \hat{\beta}_{52} & \hat{\beta}_{53} \\ \hat{\beta}_{61} & \hat{\beta}_{62} & \hat{\beta}_{63} \end{pmatrix} \begin{pmatrix} a_3 \\ b_3 \\ c_3 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$$

Ces poids sont :

$$W_1 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} & \hat{\beta}_{13} \\ \hat{\beta}_{41} & \hat{\beta}_{42} & \hat{\beta}_{43} \\ \hat{\beta}_{71} & \hat{\beta}_{72} & \hat{\beta}_{73} \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, W_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_{21} & \hat{\beta}_{22} & \hat{\beta}_{23} \\ \hat{\beta}_{41} & \hat{\beta}_{42} & \hat{\beta}_{43} \\ \hat{\beta}_{51} & \hat{\beta}_{52} & \hat{\beta}_{53} \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix},$$

$$W_3 = \begin{pmatrix} a_3 \\ b_3 \\ c_3 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{31} & \hat{\beta}_{32} & \hat{\beta}_{33} \\ \hat{\beta}_{51} & \hat{\beta}_{52} & \hat{\beta}_{53} \\ \hat{\beta}_{61} & \hat{\beta}_{62} & \hat{\beta}_{63} \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$$

On obtient une matrice de coefficients cointégrants exactement identifiés de la forme :

$$\tilde{\hat{\beta}} = \begin{pmatrix} -1 & \tilde{\hat{\beta}}_{12} & \tilde{\hat{\beta}}_{13} \\ \tilde{\hat{\beta}}_{21} & -1 & \tilde{\hat{\beta}}_{23} \\ \tilde{\hat{\beta}}_{31} & 0 & -1 \\ 0 & \tilde{\hat{\beta}}_{42} & \tilde{\hat{\beta}}_{43} \\ \tilde{\hat{\beta}}_{51} & 0 & 0 \\ \tilde{\hat{\beta}}_{61} & \tilde{\hat{\beta}}_{62} & 0 \\ 1 & \tilde{\hat{\beta}}_{72} & \tilde{\hat{\beta}}_{73} \end{pmatrix}$$

Les contraintes appliquées précédemment n'imposent aucune restriction sur l'estimation du système dynamique cointégré par maximum de vraisemblance. Il s'agit uniquement de restrictions d'identification exacte. Comme toute restriction d'identification exacte dans un modèle économique, elles ne font que révéler la manière arbitraire dont on choisit d'interpréter le modèle. On ne peut donc pas les tester statistiquement, mais on peut tester leur pertinence d'un point de vue économique, en s'assurant qu'elles donnent lieu à des relations dont les coefficients estimés sont plausibles économiquement. Or, cette plausibilité est globalement atteinte ici puisque les résultats précédents impliquent les trois relations d'équilibre suivantes :

$$primesr_t = -0{,}29\, cpr_t + 0{,}05\, depr_t + 0{,}03\, r_t - 0{,}03\, \Delta cacr_t + 1\, sinir_t$$
$$cpr_t = 4{,}64\, primesr_t - 1{,}78\, depr_t - 4{,}66\, \Delta cacr_t + 1{,}30\, sinir_t$$
$$depr_t = 0{,}48\, primesr_t + 0{,}96\, pibr_t + 0{,}10\, cpr_t - 0{,}39\, sinir_t$$



Tableau 6.4

Vecteurs de cointégration estimés normalisés et exactement identifiés

	Vecteur 1	Vecteur 2	Vecteur 3
$primesr_t$	-1.0000	4.6432	.48373
	(-)	(10.4752)	(.18715)
cpr_t	-.29218	-1.0000	.095449
	(.052699)	(-)	(.051613)
$depr_t$	.045879	-1.7860	-1.0000
	(.19062)	(5.0761)	(-)
$pibr_t$	0.00	0.00	.95807
	(-)	(-)	(.15362)
r_t	.030349	0.00	0.00
	(.010152)	(-)	(-)
$\Delta cacr_t$	-.025130	-4.6572	0.00
	(.17613)	(5.8787)	(-)
$sinir_t$	1.0000	1.3021	-.39318
	(-)	(3.4738)	(.21258)

Ces relations peuvent encore s'exprimer de la manière suivante :

$$\left(\frac{PRIMES_t}{PRIX_t}\right) = \left(\left(\frac{CP_t}{PRIX_t}\right)^{-0,29}\left(\frac{DEP_t}{PRIX_t}\right)^{0,05} e^{0,03r_t - 0,03\Delta cacr_t}\right)\left(\frac{SINI_t}{PRIX_t}\right)$$

$$\left(\frac{CP_t}{PRIX_t}\right) = \left(\frac{PRIMES_t}{PRIX_t}\right)^{4,64}\left(\frac{DEP_t}{PRIX_t}\right)^{-1,78} e^{-4,66\Delta cacr_t}\left(\frac{SINI_t}{PRIX_t}\right)^{1,30}$$

$$\left(\frac{DEP_t}{PRIX_t}\right) = \left(\frac{PRIMES_t}{PRIX_t}\right)^{0,09}\left(\frac{CP_t}{PRIX_t}\right)^{0,1}\left(\frac{SINI_t}{PRIMES_t}\right)^{-0,39}\left(\frac{PIB_t}{PRIX_t}\right)^{0,96}$$

La première relation d'équilibre indique qu'à long terme, les primes réelles sont proportionnelles aux sinistres réels. Le coefficient de proportionnalité est d'autant plus faible que le capital réel est élevé, et d'autant plus élevé que les dépenses réelles et le taux d'intérêt réel sont élevés ; il diminue si le rendement réel boursier augmente. L'élasticité unitaire des primes réelles aux sinistres réels s'explique par le fait que les primes réelles sont constituées pour l'indemnisation des sinistres, et sont donc fondamentalement déterminées par les anticipations rationnelles de ceux-ci. La relation négative des primes réelles au niveau du capital réel des assureurs semble confirmer la validité de la théorie des contraintes de capacité. La relation positive des primes réelles aux dépenses réelles s'explique par la nécessité, pour les assureurs, de percevoir les recettes nécessaires au paiement de ces charges. La relation négative des primes réelles au rendement boursier réel s'explique par le fait que l'augmentation du rendement financier des placements permet d'accepter un taux de profit technique plus faible, et donc des primes réelles plus faibles. C'est une implication du MEDAF (Modèle d'Équilibre des Actifs Financiers) de l'assurance. On devrait donc aussi observer une relation négative entre primes réelles et taux d'intérêt réel, mais la relation positive obtenue ici s'explique par la corrélation forte entre le taux d'intérêt réel et le taux de croissance réel de l'activité économique (conformément à la théorie de l'équilibre général) : une augmentation du taux d'intérêt réel reflète une augmentation de la croissance globale, qui dope la demande de produits d'assurance et pousse les primes à la hausse.

La deuxième relation d'équilibre signifie qu'à long terme, le capital réel des assureurs croît si les sinistres réels augmentent, car cela les incite à lever de nouveaux capitaux. Le capital réel est également, et très logiquement, affecté à la hausse par le niveau des primes réelles émises et affecté à la baisse par le niveau des dépenses réelles. Une augmentation du taux de rendement boursier réel permet de détenir moins de capitaux réels car le rendement financier de ceux-ci est plus élevé.

La troisième relation d'équilibre suggère que les dépenses réelles liées à la vente des produits d'assurance et à la gestion des sinistres augmentent logiquement avec le niveau des primes émises et le capital réel des assureurs, qui mesure la capacité de production du secteur. Toutefois, les dépenses totales réelles diminuent si le rapport des sinistres aux primes augmente, ce qui semble suggérer qu'une augmentation des sinistres incite les assureurs à diminuer leurs frais de gestion pour maintenir la rentabilité, et qu'une croissance des sinistres induit une croissance des dépenses de gestion inférieure à la croissance des dépenses de distribution induite par une croissance des primes émises.

Sur la base des résultats précédents, il semble raisonnable de tester des restrictions de suridentification, en plus des restrictions d'identification exacte. On vérifie $\beta_{73} = -\beta_{13}$ et $\beta_{43} = 1$, avec toutes les restrictions précédentes. On obtient les résultats suivants :

Tableau 6.5

Vecteurs de cointégration estimés normalisés avec restrictions de suridentification ($\beta_{73} = -\beta_{13}$ et $\beta_{43} = 1$)

	Vecteur 1	Vecteur	2 Vecteur 3
$primesr_t$	-1.0000	3.0642	.51845
	(-)	(5.6436)	(.20835)
cpr_t	-.28728	-1.0000	.13054
	(.047368)	(-)	(.032923)
$depr_t$	.038765	-1.6938	-1.0000
	(.16441)	(2.5250)	(-)
$pibr_t$	-.0000	.0000	1.0000
	(-)	(-)	(-)
r_t	.029035	.0000	-.0000
	(.010034)	(-)	(-)
$\Delta cacr_t$	.0085136	-3.1791	.0000
	(.13634)	(2.5978)	(-)
$sinir_t$	1.0000	1.9917	-.51845
	(-)	(2.1812)	(.20835)

Ces restrictions de suridentification ne sont pas rejetées puisque le test Chi-2 (2) du rapport de vraisemblance donne une valeur de 0.44216, à laquelle est associée une probabilité critique de 80,2 %. Le logarithme de la fonction de vraisemblance avec seulement des restrictions d'identification exacte vaut 393.7614 alors qu'il vaut, avec les restrictions de suridentification, 393.5404.

Les nouvelles relations d'équilibre peuvent s'écrire :

$$primesr_t = -0{,}29\, cpr_t + 0{,}04\, depr_t + 0{,}03\, r_t + 0{,}01\, \Delta cacr_t + 1\, sinir_t$$
$$cpr_t = 3{,}06\, primesr_t - 1{,}69\, depr_t - 3{,}18\, \Delta cacr_t + 1{,}99\, sinir_t$$
$$depr_t = 1\, pibr_t + 0{,}13\, cpr_t - 0{,}51\, (sinir_t - primesr_t)$$

Comme précédemment, on peut représenter ces relations d'équilibre de la manière suivante :

$$\left(\frac{PRIMES_t}{PRIX_t}\right) = \left(\left(\frac{CP_t}{PRIX_t}\right)^{-0,29}\left(\frac{DEP_t}{PRIX_t}\right)^{0,04} e^{0,03r_t+0,01\Delta cacr_t}\right)\left(\frac{SINI_t}{PRIX_t}\right)$$

$$\left(\frac{CP_t}{PRIX_t}\right) = \left(\frac{PRIMES_t}{PRIX_t}\right)^{3,06}\left(\frac{DEP_t}{PRIX_t}\right)^{-1,69} e^{-3,18\Delta cacr_t}\left(\frac{SINI_t}{PRIX_t}\right)^{1,99}$$

$$\left(\frac{DEP_t}{PRIX_t}\right) = \left(\left(\frac{CP_t}{PRIX_t}\right)^{0,13}\left(\frac{SINI_t}{PRIMES_t}\right)^{-0,51}\right)\left(\frac{PIB_t}{PRIX_t}\right)$$

Ces nouvelles relations d'équilibre apportent peu de changements en termes de signes des coefficients. Seul le signe du coefficient du taux de rendement réel des actions change : il est désormais le même que celui du coefficient du taux d'intérêt réel, et s'oppose à ce que suggère le MEDAF de l'assurance.

Sur la base des résultats précédents, on ajoute à ces restrictions l'hypothèse que $\beta_{23} = 0$. On obtient alors les résultats suivants :

Tableau 6.6

Vecteurs de cointégration estimés normalisés avec restrictions de suridentification ($\beta_{73} = -\beta_{13}$, $\beta_{43} = 1$ et $\beta_{23} = 0$)

	Vecteur 1	Vecteur 2	Vecteur 3
$primesr_t$	-1.0000	9.1297	.60767
	(-)	(7.7425)	(.34110)
cpr_t	-1.1772	-1.0000	.0000
	(1.8462)	(-)	(-)
$depr_t$	4.8323	-1.6442	-1.0000
	(7.3529)	(4.4203)	(-)
$pibr_t$	-.0000	-.0000	1.0000
	(-)	(-)	(-)
r_t	-.031781	.0000	-.0000
	(.20409)	(-)	(-)
$\Delta cacr_t$	.49941	.36283	.0000
	(.64560)	(-)	(-)
$sinir_t$	1.0000	-9.1574	-.60767
	(-)	(7.7777)	(.34110)

Les relations cointégrées sont donc :

$$lprimesr_t = -1{,}18\, lcpr_t + 4{,}83\, ldepr_t - 0{,}03\, oblr_t + 0{,}50\, \Delta lcacr_t + 1\, lsinr_t$$

$$lcpr_t = 9{,}12\, lprimesr_t - 1{,}64\, ldepr_t + 0{,}36\, \Delta lcacr_t - 9{,}15\, lsinr_t$$

$$ldepr_t = 1\, lpibr_t - 0{,}60\, (lsinr_t - lprimesr_t)$$

On peut les représenter ainsi :

$$\left(\frac{PRIMES_t}{PRIX_t}\right) = \left(\left(\frac{CP_t}{PRIX_t}\right)^{-1,18}\left(\frac{DEP_t}{PRIX_t}\right)^{4,83} e^{-0,03r_t+0,50\Delta cacr_t}\right)\left(\frac{SINI_t}{PRIX_t}\right)$$

$$\left(\frac{CP_t}{PRIX_t}\right) = \left(\frac{PRIMES_t}{PRIX_t}\right)^{9,12}\left(\frac{DEP_t}{PRIX_t}\right)^{-1,64} e^{0,36\Delta cacr_t}\left(\frac{SINI_t}{PRIX_t}\right)^{-9,15}$$

$$\left(\frac{DEP_t}{PRIX_t}\right) = \left(\frac{SINI_t}{PRIMES_t}\right)^{-0,6}\left(\frac{PIB_t}{PRIX_t}\right)$$

Du point de vue de leur plausibilité économique, ces résultats sont particulièrement intéressants. Cette fois, dans la relation d'équilibre des primes, le taux d'intérêt a le coefficient négatif suggéré par le MEDAF de l'assurance. En revanche, le rendement boursier réel a un coefficient positif : il joue vraisemblablement ici le rôle de « proxy » (c'est-à-dire d'une variable de remplacement) pour le taux de croissance réel de l'activité avec lequel il est très corrélé. Lorsque la croissance réelle est forte, la demande de produits d'assurance est élevée et pousse les primes réelles à la hausse, conformément à la théorie de la capacité contrainte. Avec cette nouvelle restriction, la relation d'équilibre du capital des assureurs devient alors plus aisée à interpréter : les primes émises font augmenter le capital, tandis que les sinistres et les dépenses le réduisent. Les gains boursiers provoquent également une augmentation du capital. Toutefois, les restrictions imposées ne sont pas acceptées puisqu'un test Chi-2 (3) du rapport de vraisemblance donne une valeur de 85,5719, à laquelle est associée une probabilité critique de 0,000 %.

On s'intéresse maintenant au modèle *VAR* complet, qui est à court terme, dont la solution d'équilibre (ou à long terme) est donnée par les trois vecteurs de cointégration définis précédemment.

Puisque la dernière restriction n'est pas acceptée, on conserve seulement les vecteurs de cointégration avec les hypothèses de suridentification $\beta_{73} = -\beta_{13}$ et $\beta_{43} = 1$.

On définit de nouvelles variables *ecm*1, *ecm*2 et *ecm*3, qui sont les écarts par rapport aux trois relations d'équilibre :

$$ecm1_t = -0{,}29\, cpr_t + 0{,}04\, depr_t + 0{,}03\, r_t + 0{,}01\, \Delta cacr_t + 1\, sinir_t - primesr_t$$
$$ecm2_t = 3{,}06\, primesr_t - 1{,}69\, depr_t - 3{,}18\, \Delta cacr_t + 1{,}99\, sinir_t - cpr_t$$
$$ecm3_t = 1\, pibr_t + 0{,}13\, cpr_t - 0{,}51\, (sinir_t - primesr_t) - depr_t$$

Le modèle *VAR*(2) cointégré à correction d'erreur fournit des résultats d'estimation sur la période 1966–1999, pour les sept équations simultanées du système. On obtient les résultats suivants pour l'équation relative aux primes :

Tableau 6.7

Équation de $\Delta primesr_t$

	Coefficient estimé	Écart-type	Statistique *t*	[Probabilité critique]
c	0.98187	.28464	3.4495	[.002]
$\Delta primesr_{t-1}$	-0.03514	.22111	-0.1589	[.875]
Δcpr_{t-1}	-0.09543	.08160	-1.1694	[.254]
$\Delta depr_{t-1}$	0.17098	.28779	0.5941	[.558]
$\Delta pibr_{t-1}$	-0.12923	.43508	-0.2970	[.769]
Δr_{t-1}	-0.01011	.00551	-1.8341	[.080]
$\Delta(\Delta cacr)_{t-1}$	-0.02089	.03225	-0.6476	[.524]
$\Delta sinir_{t-1}$	-0.00835	.07673	-0.1089	[.914]
$ecm1_{t-1}$	0.30047	.08171	3.6771	[.001]
$ecm2_{t-1}$	-0.00399	.01496	-0.2666	[.792]
$ecm3_{t-1}$	0.33231	.20087	1.6544	[.112]
R^2	.69601			

Dans l'équation des variations des primes réelles, les variables différenciées prennent en compte les effets à court terme des variables passées, tandis que les variables *ecm* représentent les différences par rapport aux relations à long terme. Le seul coefficient significatif (même si le caractère « significatif » n'a pas une grande importance ici, puisque le terme d'erreur n'est pas un bruit blanc) concerne les taux d'intérêt réels à long terme, dont le coefficient négatif confirme les conclusions des modèles financiers de l'assurance (comme le MEDAF) : quand les taux d'intérêt diminuent, les primes augmentent, afin de compenser la baisse des résultats financiers. Il en est de même quand l'indice boursier diminue. Cet effet est donc valable à long terme (comme montré précédemment) et à court terme, même s'il ne s'agit, pour l'instant encore, que des effets des variables retardées.

La variable $ecm1_{t-1}$, par exemple, est la différence entre la valeur d'équilibre des primes au temps t et la valeur observée des primes au temps $t-1$. Quand les primes sont trop faibles par rapport à la relation d'équilibre, elles doivent augmenter pour retrouver l'équilibre : il est donc normal que le coefficient de $ecm1_{t-1}$ soit positif.

Puisque les coefficients $ecm2_{t-1}$ et $ecm3_{t-1}$ ne sont pas significatifs, les primes ne répondent pas au déséquilibre concernant le niveau des fonds propres et les dépenses.

Après l'étude du *VAR*(2) cointégré ne contenant que les restrictions de suridentification acceptées, on examine un *VAR*(2) avec la restriction de suridentification supplémentaire donnant des vecteurs de cointégration plus plausibles économiquement. En effet, Pesaran [PES 1997] montre que le test du ratio de vraisemblance des restrictions sur les coefficients des vecteurs de cointégration a tendance à rejeter abusivement des restrictions vraies. Dès lors, il est sage d'utiliser des vecteurs de cointégration qui ont une bonne plausibilité économique, même s'ils sont le résultat de restrictions de suridentification rejetées (comme c'est le cas dans le modèle *VAR* cointégré à correction d'erreur, qui utilise les vecteurs de cointégration avec l'hypothèse supplémentaire que $\beta_{23} = 0$).

C'est la raison pour laquelle il est intéressant d'observer le *VAR* obtenu lorsqu'on impose les restrictions sur les vecteurs qui donnent des résultats économiquement vraisemblables (à $\beta_{73} = -\beta_{13}$ et $\beta_{43} = 1$, on ajoute $\beta_{23} = 0$).

Tableau 6.8

Équation de $\Delta primesr_t$

	Coefficient estimé	Écart-type	Statistique t	[Probabilité critique]
c	.756700	.30634	2.47020	[.021]
$\Delta primesr_{t-1}$	.159840	.22679	.70481	[.488]
Δcpr_{t-1}	.004936	.07107	.06946	[.945]
$\Delta depr_{t-1}$	-.083749	.32407	-.25843	[.798]
$\Delta pibr_{t-1}$	.001925	.50952	.00378	[.997]
Δr_{t-1}	.002905	.00500	.58084	[.567]
$\Delta(\Delta cacr)_{t-1}$	-.016564	.02507	-.66083	[.515]
$\Delta sinir_{t-1}$	.004676	.08970	.05214	[.959]
$ecm1_{t-1}$	-.031495	.01853	-1.69940	[.103]
$ecm2_{t-1}$	-.011736	.01777	-.66052	[.515]
$ecm3_{t-1}$	.099328	.35493	.27985	[.782]
R^2	.59960			

On définit alors de nouvelles variables *ecm*1, *ecm*2, *ecm*3, qui sont les écarts par rapport aux trois relations d'équilibre :

$$ecm1_t = -1{,}18\, cpr_t + 4{,}83\, depr_t - 0{,}03\, r_t + 0{,}50\, \Delta cacr_t + 1\, sinir_t - primesr_t$$
$$ecm2_t = 9{,}12\, primesr_t - 1{,}64\, depr_t + 0{,}36\, \Delta cacr_t - 9{,}15\, sinir_t - cpr_t$$
$$ecm3_t = pibr_t - 0{,}60\, (sinir_t - primesr_t) - depr_t$$

Puisque le terme d'erreur n'est pas un bruit blanc, les écarts types publiés sont erronés, et l'on ne peut attacher d'importance au caractère significatif ou non des coefficients. Seul le signe du coefficient des variations du rendement boursier de l'année antérieure ne change pas par rapport à l'estimation précédente. Le signe des sinistres correspond mieux à la réalité : en cas de hausse des sinistres l'année précédente, les primes augmentent.

EXERCICE 4 ASSURANCE ET MODÈLE *VAR* CONDITIONNEL À CORRECTION D'ERREUR

Énoncé

Le modèle *VAR* cointégré représente la relation des sept variables avec leur passé, sous forme de correction d'erreur autour d'une relation d'équilibre. Cette modélisation permet d'identifier des relations à long terme entre les variables, et les effets à court terme des variables passées sur les variables contemporaines. Dans ce modèle ne sont pas représentées les relations contemporaines, c'est-à-dire les relations entre les différentes variables à la même date. On souhaite à présent s'y intéresser. Pour améliorer la modélisation, il est également naturel d'envisager la formation des principales variables d'intérêt (*primesr*, *cpr*, *depr*) conditionnellement aux autres variables. L'objet de cet exercice est donc de spécifier un modèle à correction d'erreur pour *primesr*, *cpr*, et *depr*, conditionnel aux variables *pibr*, *r*, $\Delta cacr$ et *sinir*, qui permette d'introduire les incidences contemporaines (ou instantanées) de ces quatre variables sur *primesr*, *cpr* et *depr*.

Solution

Le vecteur u_t des termes d'erreur du système dynamique cointégré est supposé non autocorrélé et distribué d'après une normale de matrice de variance et de covariance Σ c'est-à-dire : $u_t \sim N(0, \Sigma)$. Il en résulte la distribution conditionnelle suivante de ΔX_t par rapport à son passé :

$$\Delta X_t \,|X_{t-1}\,, X_{t-2}, \ldots \sim N\left(\mu + \alpha\beta' X_{t-1} + A_1 \Delta X_{t-1}, \Sigma\right)$$

On partitionne le vecteur X_t en $X_t = \begin{pmatrix} Y_t \\ Z_t \end{pmatrix}$, où Y_t est un vecteur 3×1, et Z_t un vecteur 4×1 :

$$Y_t = \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix} = \begin{pmatrix} primesr_t \\ cpr_t \\ depr_t \end{pmatrix}, \quad Z_t = \begin{pmatrix} X_{4t} \\ X_{5t} \\ X_{6t} \\ X_{7t} \end{pmatrix} = \begin{pmatrix} pibr_t \\ r_t \\ \Delta cacr_t \\ sinir_t \end{pmatrix}$$

On partitionne également le vecteur μ et la matrice Σ :

$$\mu^1 = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_{3t} \end{pmatrix}, \quad \mu^2 = \begin{pmatrix} \mu_4 \\ \mu_5 \\ \mu_6 \\ \mu_7 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

où

$$\Sigma_{11} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}, \quad \Sigma_{12} = \begin{pmatrix} \sigma_{14} & \sigma_{15} & \sigma_{16} & \sigma_{17} \\ \sigma_{24} & \sigma_{25} & \sigma_{26} & \sigma_{27} \\ \sigma_{34} & \sigma_{35} & \sigma_{36} & \sigma_{37} \end{pmatrix},$$

$$\Sigma_{21} = \begin{pmatrix} \sigma_{41} & \sigma_{42} & \sigma_{43} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} \\ \sigma_{61} & \sigma_{62} & \sigma_{63} \\ \sigma_{71} & \sigma_{72} & \sigma_{73} \end{pmatrix}, \quad \Sigma_{22} = \begin{pmatrix} \sigma_{44} & \sigma_{45} & \sigma_{46} & \sigma_{47} \\ \sigma_{54} & \sigma_{55} & \sigma_{56} & \sigma_{57} \\ \sigma_{64} & \sigma_{65} & \sigma_{66} & \sigma_{67} \\ \sigma_{74} & \sigma_{75} & \sigma_{76} & \sigma_{77} \end{pmatrix}$$

On partitionne les matrices α et A_1 de manière adéquate :

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

où α_1 est une matrice 3×3 contenant les trois premières lignes de α, et α_2 une matrice 4×3 contenant les quatre dernières lignes de α.

$$A_1 = \begin{pmatrix} A_{11} \\ A_{12} \end{pmatrix}$$

où A_{11} est une matrice 3×7 avec les trois premières lignes de A, et A_{21} une matrice 4×7 avec les quatre dernières lignes de A.

L'hypothèse de normalité précédemment évoquée permet de déduire la distribution conditionnelle de ΔY_t conditionnellement à Z_t et au passé de X, et la distribution marginale de ΔZ_t :

$$\begin{aligned} &\Delta Y_t \,|Z_t, X_{t-1}, X_{t-2} \ldots \\ &\sim N \begin{pmatrix} \mu_1 + \Sigma_{12}\Sigma_{22}^{-1} (\Delta Z_t - \mu_2) + \left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right) \beta' X_{t-1} + \left(A_{11} - \Sigma_{12}\Sigma_{22}^{-1} A_{21}\right) \Delta X_{t-1} \\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix} \end{aligned}$$

$$\Delta Z_t \,|Z_{t-1}, Z_{t-2} \ldots \; \sim N \left(\mu_2 + \alpha_2 \beta' X_{t-1} + A_{21} \Delta X_{t-1}, \Sigma_{22}\right)$$

On peut encore écrire :

$$\begin{aligned} \Delta Y_t &= \left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2\right) + \left(\Sigma_{12}\Sigma_{22}^{-1}\right) \Delta Z_t + \left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right) \beta' X_{t-1} \\ &\quad + \left(A_{11} - \Sigma_{12}\Sigma_{22}^{-1} A_{21}\right) \Delta X_{t-1} + u_{1.2t} \\ \Delta Z_t &= \mu_2 + \alpha_2 \beta' X_{t-1} + A_{21} \Delta X_{t-1} + u_{2t} \end{aligned}$$

où $u_{1.2t} = u_t^1 - \Sigma_{12}\Sigma_{22}^{-1} u_t^2$.

La première équation matricielle est un modèle à correction d'erreur conditionnel en forme réduite, tandis que la deuxième est un modèle à correction d'erreur marginal [BOS 1992]. Ces deux équations sont un simple reparamétrage du système *VAR* cointégré original. Elles contiennent donc la même information, mais présentée différemment.

Ici, $\left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2\right)$ et $u_{1.2t}$ sont des vecteurs 3×1, $\Sigma_{12}\Sigma_{22}^{-1}$ est une matrice 3×4, $\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)$ une matrice 3×3, $\beta' X_{t-1}$ un vecteur 3×1, $\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right) \beta'$ une matrice 3×7, et $\left(A_{11} - \Sigma_{12}\Sigma_{22}^{-1} A_{21}\right)$ une matrice 3×7.

Le modèle conditionnel peut être estimé isolément si Z_t est faiblement exogène pour les paramètres de l'équation [1]. Si c'est le cas, on peut ignorer le modèle marginal pour réaliser de l'inférence sur les paramètres du modèle à correction d'erreur conditionnel, y compris sur les paramètres des vecteurs de cointégration. Une exogénéité faible apparaît effectivement si $\alpha_2 = 0$, comme l'a prouvé Johansen. Cette condition a pour but de rendre indépendants les deux systèmes. Elle a pour conséquence d'éliminer la présence du terme correcteur d'erreur dans le modèle marginal (qui est toutefois dans le modèle conditionnel).

La manière correcte de réaliser de l'inférence statistique sur ce modèle conditionnel dépend donc des propriétés d'exogénéité des variables.

Les variables *pibr*, *r*, Δ*cacr* et *sinir* sont faiblement exogènes pour les paramètres du modèle conditionnel si, dans le *VAR* présenté précédemment, leurs variations ne dépendent pas des écarts à l'équilibre retardés, c'est-à-dire si les coefficients α_{41}, α_{42}, α_{43}, α_{51}, α_{52}, α_{53}, α_{61}, α_{62}, α_{63}, α_{71}, α_{72} et α_{73} sont nuls simultanément.

On réalise un test de Wald afin de vérifier si les douze coefficients de la matrice α_2 (les coefficients des variables par rapport auxquelles on « conditionnalise » dans le modèle conditionnel, qui forme une partition de la matrice α) sont nuls. Le résultat est 36,0196 et suit une distribution Chi-2 à 12 degrés de liberté, mais affiche une probabilité de 0,000 % à la droite de la valeur réalisée du test, montrant que l'hypothèse $\alpha_2 = 0$ est rejetée. Les quatre variables *pibr*, *r*, Δ*cacr* et *sinir* ne sont pas faiblement exogènes.

Ainsi, au lieu d'estimer seules les trois équations du modèle conditionnel, il convient de les estimer simultanément avec les équations du *VAR* des variables par rapport auxquelles on conditionne le modèle.

Les résultats de l'estimation de modèle *VAR* à correction d'erreur conditionnel sont les suivants pour l'équation des primes :

Tableau 6.9

Équation de Δ*primesr*$_t$

	Coefficient estimé	Écart-type	Statistique *t*	[Probabilité critique]
c	47893	.33710	1.421	[.155]
ecm1(-1)	-.03191	.02036	-1.567	[.117]
ecm2(-1)	.00964	.01117	0.086	[.931]
ecm3(-1)	-.15629	.23806	-0.656	[.511]
$\Delta primesr_{t-1}$	.35077	.22523	1.557	[.119]
Δcpr_{t-1}	.00192	.04889	0.039	[.969]
$\Delta depr_{t-1}$	-.18779	.29787	-0.630	[.528]
$\Delta pibr_{t-1}$	.39280	.59185	0.664	[.507]
Δr_{t-1}	.00143	.00343	0.415	[.678]
$\Delta(\Delta cacr)_{t-1}$	-.03528	.01844	-1.914	[.056]
$\Delta sinir_{t-1}$	-.02448	.07571	-0.324	[.746]
$\Delta pibr_t$	.10156	.32699	0.311	[.756]
Δr_t	.00309	.00262	1.1785	[.239]
$\Delta(\Delta cacr)_t$	-.00499	.01443	-0.346	[.729]
$\Delta sinir_t$	.06478	.04780	1.355	[.175]
R^2=	.62791			

1. Une variable aléatoire Z_t est dite « faiblement exogène » pour un paramètre θ, si elle peut être prise comme fixe et donnée sans causer de perte d'information lors de l'inférence sur θ. On peut ajouter, à titre de remarque, que si l'on a à la fois exogénéité faible et non-causalité de Granger, alors Z_t est dite « fortement exogène ».

Parmi les quatre variables instantanées, celles du PIB, des sinistres et du rendement des actions ont le signe attendu. La croissance génère une hausse des primes (suite à l'augmentation du niveau de vie, la valeur des biens à assurer augmente, de nouvelles activités sont entreprises...), une hausse des rendements de l'indice boursier permet de diminuer les primes et une augmentation des sinistres les fait croître. Les taux d'intérêt à long terme semblent influencer positivement les primes. Toutefois, le coefficient n'est pas significatif et, dans cette équation, seule la variable relative au rendement des actions retardé est significative, avec le signe attendu.

Exercice 5 Assurance et modèle structurel à correction d'erreur

Énoncé

Après avoir estimé le modèle *VAR* à correction d'erreur conditionnel, il faut maintenant estimer un modèle structurel à correction d'erreur, pour faire apparaître les relations contemporaines simultanées entre les variables d'intérêt (*primesr*, *capr*, *depr*).

Solution

À partir du modèle à correction d'erreur conditionnel en forme réduite, on peut spécifier différents modèles structurels, en particulier un modèle structurel à équations simultanées à correction d'erreur.

On partitionne la matrice β' de la manière suivante : $\beta' = \left(\Gamma \vdots B\right)$, où Γ est supposé être une matrice non singulière 3×3, et B est une matrice 3×4, ce qui induit :

$$\beta' X_t = \Gamma Y_t + BZ_t = v_t$$

Partitionner la matrice β en deux matrices Γ et B permet de réinterpréter les relations de cointégration dans une perspective de modèle structurel reliant des variables endogènes Y à des variables exogènes Z. Le vecteur $3 \times 1 v_t$ est ainsi défini comme le vecteur des écarts aux équilibres à la période t.

Les matrices Γ et B sont identifiées moyennant des hypothèses d'identification, décrites par Park [PAR 1990] et Saikkonen [SAI 1991]. Les hypothèses d'identification exacte de β examinées précédemment valent aussi pour Γ et B, puisque ces matrices sont une partition de β et, avec la nouvelle notation de β partitionné en Γ et B, elles peuvent se réécrire de la manière suivante [(1)] :

$\Gamma_{ii} = -1, i = 1 \ldots 3$ pour les hypothèses de normalisation et $R_i \left(\Gamma_i \vdots B_i\right)' = 0, i = 1 \ldots 3$ pour les hypothèses supplémentaires d'identification exacte.

Le vecteur $\left(\Gamma_i \vdots B_i\right)'$ est la i^e colonne de β. Chaque matrice de restrictions R_i doit être d'ordre 2×7 et de rang 2 : un vecteur de cointégration particulier est exactement identifié si, en plus de la restriction de normalisation, on lui impose deux restrictions linéairement indépendantes.

Bien entendu, on peut également imposer, en plus des restrictions de normalisation et des restrictions supplémentaires d'identification exacte, des restrictions supplémentaires

1. En l'absence de restrictions interéquations, qui seraient une autre manière d'identifier les coefficients.

de suridentification. C'est le cas ici si une ou plusieurs matrices R_i ont plus de deux lignes et sont d'un ordre supérieur à 2. Seules les restrictions de suridentification peuvent être testées statistiquement.

On obtient un modèle structurel à correction d'erreur en reparamétrant le modèle à correction d'erreur conditionnel en forme réduite de la façon suivante : on substitue $\left(\Gamma \vdots B\right)$ à β' et l'on multiplie l'ensemble par une matrice non singulière 3×3 notée Γ_0 :

$$\begin{aligned}\Gamma_0 \Delta Y_t = \Gamma_0 & \left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2\right) \\ & + \Gamma_0\left(\Sigma_{12}\Sigma_{22}^{-1}\right)\Delta Z_t + \Gamma_0\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)\left(\Gamma Y_{t-1} + BZ_{t-1}\right) \\ & + \Gamma_0\left(A_{11} - \Sigma_{12}\Sigma_{22}^{-1}A_{21}\right)\Delta X_{t-1} + \Gamma_0 u_{1.2t}\end{aligned}$$

Le nouveau vecteur de termes d'erreur $\xi_t = \Gamma_0 u_{1.2t}$ a une matrice de variance et de covariance $\Xi = \Gamma_0\left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)\Gamma_0'$.

On écrit le modèle structurel sous une forme plus simple :

$$\Gamma_0 \Delta Y_t = b + B_0 \Delta Z_t + \Lambda\left(\Gamma Y_{t-1} + BZ_{t-1}\right) + \Gamma_1 \Delta Y_{t-1} + B_1 \Delta Z_{t-1} + \xi_t$$

moyennant les définitions suivantes :

$$b = \Gamma_0\left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2\right), \quad B_0 = \Gamma_0\left(\Sigma_{12}\Sigma_{22}^{-1}\right), \quad \Lambda = \Gamma_0\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)$$

où $\left(\Gamma_1 \vdots B_1\right) = \Gamma_0\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)$

Ces nouvelles notations sont des matrices dont les éléments sont notés de la manière suivante :

$$\Gamma_0 = \begin{pmatrix} \gamma_{0,11} & \gamma_{0,12} & \gamma_{0,13} \\ \gamma_{0,21} & \gamma_{0,22} & \gamma_{0,23} \\ \gamma_{0,31} & \gamma_{0,32} & \gamma_{0,33} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ b \end{pmatrix}, \quad B_0 = \begin{pmatrix} b_{0,11} & b_{0,12} & b_{0,13} \\ b_{0,21} & b_{0,22} & b_{0,23} \\ b_{0,31} & b_{0,32} & b_{0,33} \end{pmatrix},$$

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix}$$

$$\Gamma_1 = \begin{pmatrix} \gamma_{11,1} & \gamma_{12,1} & \gamma_{13,1} \\ \gamma_{21,1} & \gamma_{22,1} & \gamma_{23,1} \\ \gamma_{31,1} & \gamma_{32,1} & \gamma_{33,1} \end{pmatrix}, \quad B_1 = \begin{pmatrix} b_{11,1} & b_{12,1} & b_{13,1} & b_{14,1} \\ b_{21,1} & b_{22,1} & b_{23,1} & b_{24,1} \\ b_{31,1} & b_{32,1} & b_{33,1} & b_{34,1} \end{pmatrix}$$

Le problème de l'identification de Γ et B ayant déjà été traité, il reste à aborder celui de l'identification des paramètres restants : $\left(\Gamma_1 \vdots B_1\right) = \Gamma_0\left(\alpha_1 - \Sigma_{12}\Sigma_{22}^{-1}\alpha_2\right)$, qui décrivent les ajustements à court terme, par rapport à Γ et à B (ceux-ci étant les paramètres des relations d'équilibre à long terme). À nouveau, on impose des restrictions de normalisation et des restrictions supplémentaires d'identification, inspirées si possible par la théorie économique. Les restrictions de normalisation peuvent s'exprimer de la manière suivante :

$$\Gamma_{ii} = 1, \quad i = 1 \ldots 3$$

La sélection de restrictions supplémentaires d'identification sur les paramètres à court terme est difficile, car les propositions dégagées par la théorie économique portent essentiellement sur les relations d'équilibre, et expliquent généralement peu les ajustements à court terme.

Il serait donc raisonnable de commencer par des restrictions d'identification exacte théoriquement plausibles, dans la perspective d'appliquer éventuellement ensuite, au vu des résultats, des restrictions supplémentaires de suridentification « testables ». On met en oeuvre ici la méthode proposée initialement par Boswijk [BOS 1992], qui suggère de faire les restrictions d'identification exacte qui contraignent la matrice Λ à être diagonale :

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ 0 & \lambda_{22} & 0 \\ 0 & 0 & \lambda_{33} \end{pmatrix}$$

Pour chacune des trois équations, on impose donc deux restrictions sur les coefficients, en l'occurrence des restrictions de nullité. Avec la restriction de normalisation, on dispose donc de trois restrictions, qui permettent l'identification exacte.

Avec toutes ces restrictions d'identification exacte, le modèle structurel à correction d'erreur à estimer est donc :

$$\begin{aligned}
\Delta primesr_t = {} & b_1 - \gamma_{0,12}\Delta cpr_t - \gamma_{0,13}\Delta depr_t \\
& + b_{0,11}\Delta pibr_t + b_{0,12}\Delta oblr_t + b_{0,13}\Delta^2 cacr_t + b_{0,14} sinir_t \\
& + \lambda_1\left(\beta_{21} cpr_{t-1} + \beta_{31} depr_{t-1} + \beta_{51} oblr_{t-1} + \beta_{61}\Delta cacr_{t-1} + sinir_{t-1} - primesr_{t-1}\right) \\
& + \gamma_{11,1}\Delta primesr_{t-1} + \gamma_{12,1}\Delta cpr_{t-1} + \gamma_{13,1}\Delta depr_{t-1} \\
& + b_{11,1}\Delta pibr_{t-1} + b_{12,1}\Delta oblr_t + b_{13,1}\Delta^2 cacr_{t-1} + b_{14,1}\Delta sinir_{t-1} + \xi_{1t} \\
\Delta cpr_t = {} & b_2 - \gamma_{0,21}\Delta primesr_t - \gamma_{0,23}\Delta depr_t \\
& + b_{0,21}\Delta pibr_t + b_{0,22}\Delta oblr_t + b_{0,23}\Delta^2 cacr_t + b_{0,24} sinir_t \\
& + \lambda_2\left(\beta_{12} primesr_{t-1} + \beta_{32} depr_{t-1} + \beta_{62}\Delta cacr_{t-1} + \beta_{72} sinir_{t-1} - cpr_{t-1}\right) \\
& + \gamma_{21,1}\Delta primesr_{t-1} + \gamma_{22,1}\Delta cpr_{t-1} + \gamma_{23,1}\Delta depr_{t-1} \\
& + b_{21,1}\Delta pibr_{t-1} + b_{22,1}\Delta oblr_t + b_{23,1}\Delta^2 cacr_{t-1} + b_{24,1}\Delta sinir_{t-1} + \xi_{1t} \\
\Delta depr_t = {} & b_3 - \gamma_{0,31}\Delta primesr_t - \gamma_{0,32}\Delta cpr_t \\
& + b_{0,31}\Delta pibr_t + b_{0,32}\Delta oblr_t + b_{0,33}\Delta^2 cacr_t + b_{0,34} sinir_t \\
& + \lambda_3\left(\beta_{23} cpr_{t-1} + \beta_{33} depr_{t-1} + \beta_{43} pibr_{t-1} + \beta_{73} sinir_{t-1} - depr_{t-1}\right) \\
& + \gamma_{31,1}\Delta primesr_{t-1} + \gamma_{32,1}\Delta cpr_{t-1} + \gamma_{33,1}\Delta depr_{t-1} \\
& + b_{31,1}\Delta pibr_{t-1} + b_{32,1}\Delta oblr_t + b_{33,1}\Delta^2 cacr_{t-1} + b_{34,1}\Delta sinir_{t-1} + \xi_{3t}
\end{aligned}$$

Si l'on estime des équations exprimant des variables X conditionnellement à des variables L, ces dernières sont faiblement exogènes pour les paramètres de ces équations conditionnelles (modèle *VAR* conditionnel structurel) s'ils sont libres en variation vis-à-vis des paramètres des équations du processus de L. Dans ce cas, on ne perd pas d'information sur les paramètres des équations conditionnelles des variables X en négligeant les équations du processus des variables L.

Pour les données utilisées ici, puisque les variables *pibr*, *r*, Δ*cacr*, *sinir* ne sont pas faiblement exogènes, on estime les trois équations conditionnelles, non pas seules, mais simultanément avec les équations du *VAR* des variables par rapport auxquelles on conditionne.

Il faut aussi tenir compte du problème de l'identification des paramètres structurels. Pour qu'un système soit identifié, le nombre de variables prédéterminées (exogènes ou bien endogènes retardées) exclues de chaque équation structurelle doit être au moins égal au nombre de variables endogènes présentes dans l'équation structurelle considérée, moins une.

On estime ici un modèle conditionnel structurel identifié par maximum de vraisemblance, tenant compte de l'absence d'exogénéité faible des quatre variables *pibr*, *r*, Δ*cacr* et *sinir*. Les résultats pour l'équation des primes sont les suivants :

Tableau 6.10

Équation de $\Delta primesr_t$

Variable	Coefficient estimé	Écart-type	Statistique *t*	[Probabilité critique]
c	0.94608	0.93881	1.008	0.325
$\Delta primesr_{t-1}$	-0.32090	0.77698	-0.413	0.684
$\Delta depr_{t-1}$	0.05581	0.93699	0.060	0.953
Δr_{t-1}	-0.02593	0.02266	-1.144	0.265
$\Delta sinir_{t-1}$	-0.16256	0.27011	-0.602	0.553
ecm1(-1)	-0.50207	0.53247	-0.943	0.356
Δcpr_t	-0.48800	0.66052	-0.739	0.468
$\Delta depr_t$	1.47850	1.95830	0.755	0.458
Δr_t	0.00264	0.06627	0.040	0.969
$\Delta(\Delta cacr)_t$	-0.15949	0.39115	-0.408	0.687
$\Delta sinir_t$	-0.15974	1.76820	-0.090	0.929

Dans ce modèle structurel conditionnel, une hausse des rendements boursiers permet de diminuer le montant des primes, alors qu'une hausse des dépenses l'accroît. On retrouve également un résultat fondamental, qui va dans le même sens que la relation à long terme : une baisse des capitaux propres entraîne une hausse des primes (capacité contrainte). Toutefois, les coefficients sont peu significatifs.



Références bibliographiques

[BIE 1997] H.J. Bierens, Nonparametric Cointegration Analysis, *Journal of Econometrics*, 77, p. 379–404, 1997.

[BOS 1992] H.P. Boswijk, *Cointegration, Identification and Exogeneity*, Thesis Publishers H.P., Amsterdam, 1992.

[BOX 1970] G.E.P. Box et D.A. Pierce, Distribution of Residual Atocorremations in Autoregressive-Integrated-Movong Average Time serie Models, *Journal of American Statistical Assoiciation*, 65, p. 1509–1526, 1970.

[DAV 1993] R. Davidson, J.G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993.

[DOO 1997] J.A. Doornik, et D.F. Hendry, Modelling Dynamic Systems Using PCFIML 9.0, Thomson, 1997.

[DOR 1997] É. Dor, B. Vander Linden, M. Lopez Novella, Labour Market Policies and Unemployment Outflows, *Oxford Bulletin of Economics and Statistics*, 1997.

[ENG 1991] R.F. Engle, C.W.J. Granger, *Long-Run Economic Relationships : Readings in Cointegration*, Oxford University Press, Oxford, 1991.

[ENG 1987] R.F. Engle, C.W.J. Granger, Co-integration and Error Correction : Representation, Estimation and Testing, *Econometrica*, 55, p. 1–87, 1987.

[GRA 1974] C.W.J. Granger, P. Newbold, Spurious Regressions in Econometrics, *Journal of Econometrics*, 2, p. 111–120, 1974

[GRE 1994] A.W. Gregory, Testing for Cointegration in Linear Quadratic Models, *Journal of Business and Economic Statistics*, p. 347–360, 1994.

Hosking [HOS 1980], The Multivariate Portmanteau Statistic, *Journal of the American Statistical Association*, 75, 1980.

[JOH 1994] S. Johansen, The Role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Variables, *EconometricReviews*, 13(2), 1994.

[JOH 1991] S. Johansen, Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models, *Econometrica*, 59, p. 1551–1580, 1991.

[JOH 1990] S. Johansen, K. Juselius, Maximum Likelihood Estimation and Inference on Cointegration, with Applications to the Demand for Money, *Oxford Bulletin of Economics and Statistics*, 52, p. 169–210, 1990.

[JOH 1988] S. Johansen, Statistical Analysis of Cointegrating Vectors, *Journal of Economic Dynamics and Control*, 12, p. 231–254, 1988.

[LJU 1978] G.M. Ljung et G.E.P. Box, On a Measure of Lack of Fit in Time Series Models, Biometrika, 65, p. 297–303, 1978.

[OST 1992] M. Osterwald-Lenum, A Note with Quntiles for the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistic, *Oxford Bulletin of Economics and Statistics*, p. 461–471, 1992.

[PAR 1990] Park, Maximum Likelihood Estimation of Simultaneaous Cointegration Models, Memo, University of Aarh, 1990.

[PES 1997] M.H. Pesaran, B. Pesaran, *Working with Microfit 4.0 : An Interactive Econometric Analysis*, Oxford University Press, Oxford, 1997.

[PES 1996] M.H. Pesaran, Y. Shin, R. Smith, *Structural Analysis of Vector Error Correction Models with Exogenous I(1) variables*, University of Cambridge, 1996.

[PHI 1990a] P.C.B. Phillips, B.E. Hansen, Statistical Inference in Instrumental Variables Regression with I(1) Processes, *Review of Economic Studies*, 57, p. 99–125, 1990.

[PHI 1990b] P.C.B. Phillips, S. Ouliaris, Asymptotic Properties of Residual Based Tests for Cointegration, *Econometrica*, 578, p. 165–193, 1990.

[PHI 1986] P.C.B. Phillips, Understanding Spurious Regressions in Econometrics, *Journal of Econometrics*, 33, p. 40–311, 1986.

[SAI 1991] P. Saikkonen, Estimation of Cointegration Vectors in a Simultaneous Equation Framework, manuscript, University of Helsinky, 1991.

[WAT 1994] M.W. Watson, Vector Autoregression and Cointegration, dans *Handbook of Econometrics*, eds R.F. Engle and D. MacFadden, North Holland, p. 843–915, 1994.

[YUL 1926] G.U. Yule, Why do We Sometimes Get Nonsense-correlations between Time Series? A Study in Sampling and the Nature of Time-Series, *Journal of the Royal Statistical Society*, p. 1–64, 1926.



Variables dépendantes discrètes et volatilité conditionnelle autorégressive

Les modèles à variables dépendantes discrètes et à volatilité conditionnelle autorégressive

Ce chapitre traite de deux types de modélisation développés dans le but d'analyser des problèmes bien particuliers : d'une part les modèles qui permettent d'expliquer des choix entre plusieurs options ; d'autre part, les modèles développés spécifiquement pour représenter la dynamique de la volativité des séries financières. Dans la première partie de ce chapitre, la variable à expliquer est donc discrète ; elle ne peut prendre que quelques valeurs seulement, correspondant aux différentes options. La deuxième partie du chapitre traite des modèles à volatilité variable, particulièrement adaptés à la modélisation des rendements des actifs financiers.

1 Les modèles à variables dépendantes discrètes

1.1 Choix entre plusieurs options

Chaque individu i doit faire un choix entre J options différentes ($j = 1, 2 \ldots J$). Pour chaque individu i, et pour chaque option j, on dispose d'un vecteur w_{ij} des caractéristiques de l'option j telles qu'elles sont perçues par l'individu i. Dans les cas particuliers où les caractéristiques perçues de l'option j sont identiques, quel que soit l'individu i, l'indice i devient inutile : $w_{ij} = w_j$ pour tout i.

Pour chaque individu i, on dispose également d'un vecteur l_i des caractéristiques propres à cet individu.

Pour un individu particulier i et une option particulière j, on dispose donc d'un vecteur de caractéristiques observées z_{ij} tel que $z'_{ij} = (w'_{ij} l'_i)'$. Dans certains cas, ce vecteur se ramène simplement à $z'_{ij} = (w'_j l'_i)$.

L'utilité qu'un individu i retire de l'option j est notée U_{ij}. On suppose que cette utilité dépend des caractéristiques rassemblées dans z_{ij} et d'un facteur aléatoire (ce dernier tenant compte du fait qu'il est impossible d'identifier et d'observer tous les facteurs qui influent sur les choix d'un individu) :

$$U_{ij} = \beta' z_{ij} + \varepsilon_{ij} \tag{7.1}$$

où ε_{ij} est une composante aléatoire, et donc une variable aléatoire, qui n'est pas observable. Son espérance est supposée nulle. Le vecteur β rassemble les coefficients qui mesurent l'impact de chaque variable de z_{ij} sur l'utilité U_{ij}.

L'individu i choisit j quand l'utilité qu'il retire de l'option j est supérieure aux utilités qu'il retire de chacune des autres options :

Dire que « le choix de i est l'option j » équivaut à dire :

- $U_{ij} > U_{ik}$ pour tout $k \neq j$
- ou encore, $\beta' z_{ij} + \varepsilon_{ij} > \beta' z_{ik} + \varepsilon_{ik}$ pour tout $k \neq j$
- ou encore, $\varepsilon_{ik} - \varepsilon_{ij} < \beta' z_{ij} - \beta' z_{ik}$ pour tout $k \neq j$

La probabilité P_{ij} que le choix de l'individu i se porte sur l'option j est donc la probabilité que l'utilité qu'il retire de j soit supérieure à l'utilité qu'il retire de n'importe quelle autre option :

$$P_{ij} = P(\varepsilon_{ik} - \varepsilon_{ij} < \beta' z_{ij} - \beta' z_{ik} \quad \text{pour tout } k \neq j)$$

ou encore :

$$P_{ij} = P(\varepsilon_{ik} - \varepsilon_{ij} < \beta'(z_{ij} - z_{ik}) \quad \text{pour tout } k \neq j). \tag{7.2}$$

Dans le cas particulier où le nombre d'options s'élève à deux ($J = 2$), c'est-à-dire quand les choix d'un individu i ne peuvent porter que sur $j = 1$ ou $j = 2$, la probabilité P_{i1} que le choix de l'individu i se porte sur l'option 1 est la probabilité que l'utilité qu'il en retire soit supérieure à l'utilité qu'il retire de l'option 2 :

$$P_{i1} = P\big(\varepsilon_{i2} - \varepsilon_{i1} < \beta'(z_{i1} - z_{i2})\big)$$

ou encore

$$P_{i1} = P(\varepsilon_{i2} - \varepsilon_{i1} < \beta' Z_{i1}) = G(\beta' Z_{i1}) = G(\beta' Z_i) \tag{7.3}$$



où Z_{i1} est une manière de redéfinir les caractéristiques conjointes de l'option 1 et de l'individu i, en les définissant par rapport à celles de l'option 2 :

$$Z_{i1} = Z_i = z_{i1} - z_{i2} \tag{7.4}$$

et où $G(.)$ est la fonction de distribution cumulée (ou fonction de répartition) de la variable aléatoire $\varepsilon_{i2} - \varepsilon_{i1}$:

$$G(x) = \mathrm{P}(\varepsilon_{i2} - \varepsilon_{i1} < x) \tag{7.5}$$

1.2 Modèle Probit

Quand le nombre d'options s'élève à deux ($J = 2$) et si la variable aléatoire $\varepsilon_{i2} - \varepsilon_{i1}$ suit une distribution normale, $G(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} (e)^{-\frac{p^2}{2}} dp = \Phi(x)$. La probabilité que l'individu i fasse le choix de l'option 1 (au lieu de la 2) s'écrit alors :

$$\mathrm{P}_{i1} = \Phi(\beta' Z_{i1}) = \int_{-\infty}^{\beta' Z_{i1}} (2\pi)^{-\frac{1}{2}} (e)^{-\frac{p^2}{2}} \beta dp \tag{7.6}$$

Cette probabilité est une fonction bien définie des caractéristiques observées rassemblées dans le vecteur Z_{i1}. Les coefficients β de cette fonction sont bien sûr inconnus et inobservables, mais on peut les estimer par une technique économétrique appropriée, pourvu que l'on dispose d'un échantillon d'observations du choix et des caractéristiques d'un ensemble d'individus différents. Une fois que les coefficients du vecteur β sont estimés par des valeurs estimées $\hat{\beta}$, on peut alors appliquer la fonction $\Phi\left(\hat{\beta}' Z_{i1}\right)$ pour estimer à l'avance la probabilité qu'un nouvel individu i quelconque fasse le choix de l'option 1, en fonction des valeurs observées de ses caractéristiques Z_{i1}.

Les estimateurs $\hat{\beta}$ sont les valeurs de β qui maximisent la fonction de vraisemblance $l(\beta)$ définie par :

$$l(\beta) = \sum_{i=1}^{n} y_i \ln\left(\Phi\left(\beta' Z_i\right)\right) + \sum_{i=1}^{n} \left(1 - y_i\right) \ln\left(1 - \Phi\left(\beta' Z_i\right)\right) \tag{7.7}$$

où $y_i = 1$ quand l'individu i fait le choix de l'option 1, et $y_i = 0$ quand il choisit la 2.

Une estimation de la matrice de variance et de covariance de $\hat{\beta}$ est fournie par :

$$\hat{V}\left(\hat{\beta}\right) = \left(-E\left(\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta'}\right)\right)^{-1}\Bigg|_{\beta=\hat{\beta}} = \sum_{i=1}^{n} \left(\frac{\left(\phi\left(\hat{\beta}' Z_i\right)\right)^2}{\Phi\left(\hat{\beta}' Z_i\right)\left(1 - \Phi\left(\hat{\beta}' Z_i\right)\right)}\right) Z_i Z_i' \tag{7.8}$$

On prouve mathématiquement, sans difficulté, l'égalité de droite.

Une fois l'estimation réalisée, on définit des valeurs calculées $\hat{y}_i$ de la façon suivante :

$$\hat{y}_i = 1 \quad \text{quand} \quad \Phi\left(\hat{\beta}_i' Z_i\right) \geqslant 0{,}5$$
$$\hat{y}_i = 0 \quad \text{quand} \quad \Phi\left(\hat{\beta}_i' Z_i\right) < 0{,}5$$

Pour un individu particulier, la valeur $\hat{y}_i$ représente donc son choix « tel qu'il est impliqué par le modèle », que l'on peut comparer avec son vrai choix observé y_i.

1.3 Modèle Logit

Quand le nombre d'options s'élève à deux ($J = 2$) et si la variable aléatoire $\varepsilon_{i2} - \varepsilon_{i1}$ suit une distribution logistique, $G(x) = \dfrac{1}{1+(e)^{-x}} = \dfrac{(e)^x}{1+(e)^x}$. La probabilité que l'individu i fasse le choix de l'option 1 (au lieu de la 2) s'écrit alors :

$$\mathrm{P}_{i1} = \frac{1}{1+(e)^{-\beta' Z_{i1}}} = \frac{(e)^{\beta' Z_{i1}}}{1+(e)^{\beta' Z_{i1}}} = \Lambda(\beta' Z_{i1}) \tag{7.9}$$

Cette probabilité est une fonction bien définie des caractéristiques observées rassemblées dans le vecteur Z_{i1}. Les coefficients β de cette fonction sont bien sûr inconnus et inobservables, mais on peut les estimer par une technique économétrique appropriée, pourvu que l'on dispose d'un échantillon d'observations du choix et des caractéristiques d'un ensemble d'individus différents. Une fois que les coefficients du vecteur β sont estimés par des valeurs estimées $\hat{\beta}$, on peut alors appliquer la fonction $\Lambda\left(\hat{\beta}' Z_{i1}\right)$ pour estimer à l'avance la probabilité qu'un nouvel individu i quelconque fasse le choix de l'option 1, en fonction des valeurs observées de ses caractéristiques Z_{i1}.

Les estimateurs $\hat{\beta}$ sont les valeurs de β qui maximisent la fonction de vraisemblance $l(\beta)$ définie par :

$$l(\beta) = \sum_{i=1}^{n} y_i \ln\left(\Lambda\left(\beta' Z_i\right)\right) + \sum_{i=1}^{n}\left(1-y_i\right)\ln\left(1-\Lambda\left(\beta' Z_i\right)\right) \tag{7.10}$$

où $y_i = 1$ quand l'individu i fait le choix de l'option 1, et $y_i = 0$ quand il fait le choix de la 2.

Une estimation de la matrice de variance et de covariance de $\hat{\beta}$ est fournie par :

$$\hat{V}\left(\hat{\beta}\right) = \left(-E\left(\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta'}\right)\right)^{-1}\Bigg|_{\beta=\hat{\beta}} = \sum_{i=1}^{n}\left(\Lambda\left(\hat{\beta}' Z_i\right)\left(1-\Lambda\left(\hat{\beta}' Z_i\right)\right)\right) Z_i Z_i' \tag{7.11}$$

Après avoir estimé le modèle, on définit des valeurs calculées $\hat{y}_i$ de la manière suivante :

$$\hat{y}_i = 1 \quad \text{quand} \quad \Lambda\left(\hat{\beta}_i' Z_i\right) \geqslant 0{,}5$$

$$\hat{y}_i = 0 \quad \text{quand} \quad \Lambda\left(\hat{\beta}_i' Z_i\right) < 0{,}5$$

Pour un individu particulier, la valeur $\hat{y}_i$ représente donc son choix « tel qu'il est impliqué par le modèle », que l'on peut le comparer avec son vrai choix observé y_i.

1.4 Diagnostics sur la qualité d'ajustement des modèles Logit et Probit

Que l'on estime un modèle Probit ou Logit, on peut obtenir quelques résultats et tests à interpréter, sachant que la disponibilité de ces statistiques s'avère très variable d'un logiciel à l'autre :

- La valeur maximisée de la fonction de vraisemblance : $l\left(\hat{\beta}\right)$.
- Le critère d'Akaike : $l\left(\hat{\beta}\right) - k$.



- Le critère bayesien de Schwarz : $l\left(\hat{\beta}\right) - \frac{k}{2}\ln(n)$.
- Le critère de Hannan-Quinn : $l\left(\hat{\beta}\right) - k\ln\big(\ln(n)\big)$.

Ces critères permettent de comparer des modèles différents pour expliquer des choix y_i identiques, et de sélectionner celui pour lequel le critère est le plus élevé.

- La statistique de Pesaran-Timmermann : $PTTEST(y, \hat{y})$.

  Dans cette fonction, y est le vecteur de tous les y_i et $\hat{y}$ est le vecteur de tous les $\hat{y}_i$. $PTTEST$ est un test non paramétrique d'association entre deux variables aléatoires. Si ces deux variables sont indépendantes, $PTTEST$ suit une $N(0, 1)$ pour de grands échantillons. Par conséquent, après l'estimation d'un modèle Probit ou Logit, cette statistique permet de tester l'hypothèse que le modèle n'explique absolument pas les choix observés, c'est-à-dire que les choix prévus par le modèle n'ont aucun lien avec les choix réels observés. Quand on rejette cette hypothèse, cela signifie que le modèle arrive à expliquer au moins partiellement les choix observés.

- La mesure de qualité de l'ajustement statistique : $\sum_{i=1}^{n} \frac{\text{sign}\left(\hat{y}_i y_i\right)}{n}$.

  La fonction $\text{sign}(x)$ vaut 1 quand $x > 0$ et vaut 0 quand $x \leqslant 0$. Pour un individu i particulier, la fonction $\text{sign}\left(\hat{y}_i y_i\right)$ vaut donc 1 quand le choix $\hat{y}_i$ prévu par le modèle est correct, c'est-à-dire égal au vrai choix y_i. Dans le cas contraire, $\text{sign}\left(\hat{y}_i y_i\right) = 0$. La mesure de qualité de l'ajustement statistique représente donc le pourcentage d'individus pour lesquels le modèle prévoit correctement le choix.

- Le pseudo-R^2, qui mesure la qualité de l'ajustement statistique : $1 - l(\hat{\beta})/l(\beta_0)$.

  La fonction $l(\beta_0)$ représente la valeur maximisée de la fonction de vraisemblance quand on contraint tous les coefficients à être nuls, sauf la constante. Le pseudo-R^2 est compris entre 0 et 1 et est d'autant plus proche de 1 que le modèle est explicatif.

- La statistique Chi-2 : $2\big(l(\hat{\beta}) - l(\beta_0)\big)$.

  Si les coefficients de toutes les variables explicatives sont nuls, cette statistique suit asymptotiquement une Chi-2 à $k - 1$ degrés de liberté si $\beta' Z_i$ contient une constante et à k degrés de liberté dans le cas contraire.

1.5 Modèle Logit multinomial

Ce modèle décrit le choix entre J options différentes en supposant que les ε_{ik}, pour $k = 1 \ldots J$, suivent des lois de Weibull indépendantes. Il en résulte que la probabilité de choisir l'option j est donnée par :

$$\mathrm{P}_{ij} = \frac{e^{\beta' Z_{ij}}}{\sum_{k=1}^{J}\left(1 + e^{\beta' Z_{ik}}\right)} \tag{7.12}$$

Lorsque $J = 2$, le modèle Logit multinomial équivaut au modèle Logit. La version multinomiale présentent des limites clairement identifiables : ce modèle implique que le rapport entre les probabilités de choisir deux options différentes est indépendant de l'existence d'autres options. Si J augmente suite à l'ajout de nouvelles options, les probabilités de choix des options initiales changent, mais restent dans un même rapport. Cette propriété semble irréaliste dans beaucoup de problèmes réels, et limite donc les opportunités d'utilisation du modèle.

2 Les modèles à volatilité conditionnelle autorégressive

Les modèles à volatilité variable sont particulièrement adaptés à la modélisation des rendements des actifs financiers. Lorsqu'on observe une série temporelle financière de haute fréquence (variation du taux de change, du cours boursier...), on remarque généralement que sa volatilité, représentée statistiquement par la variance, n'est pas constante dans le temps, en d'autres termes, que sa variance change dans le temps. Le processus sous-jacent révèle donc une hétéroscédasticité [1]. En général, on constate que des périodes assez longues où la volatilité est faible sont suivies de périodes assez longues où elle est forte. En fait, des dates où la volatilité est faible ont tendance à se succéder, et des dates où elle est forte ont également tendance à se suivre. On dit également que la volatilité, ou la variance du processus, est autorégressive : elle dépend de ses valeurs passées. Les processus $ARCH(p)$ et $GARCH(p,q)$ sont définis de manière à pouvoir représenter ce type de série observée.

2.1 Les processus $ARCH(p)$

Un processus $ARCH(p)$ est un processus à hétéroscédasticité conditionnelle autorégressive d'ordre p.

Un processus stochastique $\{X_t\}$ est un processus $ARCH(p)$ quand :

$$X_t = \mu + \varepsilon_t$$

$$\text{où } f_{\varepsilon_t|X_{t-1},X_{t-2},\ldots} \sim N(0, h_t)$$

$$\text{avec } h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \quad \text{sous la condition } \alpha_i \geqslant 0 \text{ pour tout } i$$

Ce processus $\{X_t\}$ est donc tel que :

$$f_{X_t|X_{t-1},X_{t-2},\ldots} \sim N(\mu, h_t)$$

$$\text{avec } h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i (X_{t-i} - \mu)^2 \quad \text{sous la condition } \alpha_i \geqslant 0 \text{ pour tout } i$$

La distribution conditionnelle de X_t, conditionnellement à ses valeurs passées, a ainsi une variance qui dépend des valeurs passées de X_t.

L'espérance conditionnelle de X_t est fournie par :

$$E(X_t | X_{t-1}, X_{t-2}, \ldots) = \mu \quad \forall t$$

La variance conditionnelle de X_t est fournie par :

$$V(X_t | X_{t-1}, X_{t-2}, \ldots) = h_t \quad \forall t$$

1. Le chapitre 7 traite de l'hétéroscédasticité (variation de la variance dans le temps) d'un type particulier de processus stochastique : le terme d'erreur inobservable d'un modèle linéaire. Dans ce chapitre, on s'intéresse à l'hétéroscédasticité d'un processus stochastique observable.

La distribution conditionnelle de X_t à son passé est normale :

$$f_{X_t|X_{t-1},X_{t-2},\ldots} \sim N(\mu, h_t)$$

Mais la distribution non conditionnelle, ou marginale, de X_t ne l'est pas :

$$f_{X_t} \neq N$$

L'espérance non conditionnelle de X_t est fournie par :

$$E(X_t) = \mu$$

La variance non conditionnelle de X_t est fournie par :

$$V(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i}$$

quand toutes les racines du polynôme en L, $\sum_{i=1}^{p} \alpha_i L^{p-i}$, sont en dehors du cercle unité.
Remarque : le processus X_t peut s'écrire ainsi :

$$X_t = \mu + W_t U_t \quad \text{où} \quad U_t \sim N(0,1) \quad \forall t$$

et est indépendant d'une période à l'autre :

$$W_t = \left\{ \alpha_0 + \sum_{i=1}^{p} \alpha_i (X_{t-i} - \mu)^2 \right\}^{\frac{1}{2}}$$

Exemple (Processus** $ARCH(1)$**)

$$X_t = \mu + \varepsilon_t$$

où $f_{\varepsilon_t|X_{t-1},X_{t-2},\ldots} \sim N(0, h_t)$ avec $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$
et donc :

$$f_{X_t|X_{t-1},X_{t-2},\ldots} \sim N(\mu, h_t)$$

avec $h_t = \alpha_0 + \alpha_1 (X_{t-1} - \mu)^2$

2.2 Les processus $GARCH(p, q)$

Il s'agit d'une généralisation d'un processus $ARCH(p)$:

$$X_t = \mu + \varepsilon_t$$

où $f_{\varepsilon_t|X_{t-1},X_{t-2},\ldots} \sim N(0, h_t)$ avec $h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}$ sous les conditions $\alpha_i \geqslant 0$ et $\beta_j \geqslant 0$ pour tout i.
X_t est donc un processus tel que :

$$f_{X_t|X_{t-1},X_{t-2},\ldots} \sim N(\mu, h_t)$$

avec $h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i (X_{t-i} - \mu)^2 + \sum_{j=1}^{q} \beta_j h_{t-j}$

On montre que cela implique que ε_t^2 suit un processus $ARMA(m, q)$ avec $m = \max(p, q)$:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i)\, \varepsilon_{t-i}^2 + \upsilon_t - \sum_{j=1}^{q} \beta_j \upsilon_{t-j} \quad \text{où} \quad \upsilon_t = \varepsilon_t^2 - h_t$$

Exemple (Processus $GARCH(1,1)$)

$$X_t = \mu + \varepsilon_t$$

où $f_{\varepsilon_t | X_{t-1}, X_{t-2}, \ldots} \sim N(0, h_t)$ avec $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$ et donc :

$$f_{X_t | X_{t-1}, X_{t-2}, \ldots} \sim N(\mu, h_t)$$

avec $h_t = \alpha_0 + \alpha_1 (X_{t-1} - \mu)^2 + \beta_1 h_{t-1}$

Exemple (Cas particulier des processus $GARCH(1,1)$ ou processus $IGARCH(1,1)$)

Lorsque $\alpha_0 = 0$, $\alpha_1 + \beta_1 = 1$, on parle de $GARCH$ intégré ou $IGARCH(1,1)$:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + (1 - \alpha_1)\, h_{t-1}$$

ce qui implique :

$$E(h_{t+\theta} \,|\varepsilon_t, \varepsilon_{t-1}, \ldots) = h_{t+1} \quad \forall \theta \geqslant 1$$

La variance conditionnelle h_t se caractérise par une racine unitaire.

Résumé

Les modèles présentés dans la première partie de ce chapitre servent à expliquer les choix des individus entre plusieurs options. Si les choix se limitent à deux options, on utilise les modèles Probit ou Logit. Pour expliquer des choix entre un plus grand nombre d'options, on peut utiliser le modèle Logit multinomial. Ces modèles expliquent la probabilité de choisir une option en fonction d'une série de variables représentant des propriétés du décideur ou des options en présence. On trouve des informations complémentaires utiles sur ces modèles chez McFadden (1976). Les modèles à volatilité conditionnelle autorégressive permettent quant à eux de représenter l'évolution de nombreuses séries financières. Ces séries se caractérisent par une autocorrélation significative de leur volatilité alors que leur niveau ne semble pas autocorrélé. Pour des informations supplémenatires sur ces modèles il est intéressant de lire Bollerslev, Chou et Kroner [BOL 1992], Pederzoli [PED 2003] et Taylor [TAY 1986] [TAY 1994].

Problèmes et exercices

Exercice 1 Prendre la décision d'utiliser des engrais avec TSP

Énoncé

On dispose de données concernant des agriculteurs : une partie d'entre eux a choisi d'utiliser des engrais et une autre partie a choisi de ne pas en utiliser. Ces données sont dans le fichier ENGRAIS.xls, téléchargeable sur le site www.pearsoneducation.fr, et lisible par TSP.

Pour chacun de ces agriculteurs, on définit les variables suivantes :

- *ENGRAIS* vaut 1 en cas d'utilisation d'engrais et 0 sinon.
- *DETTE* mesure le montant du crédit (par hectare).
- *DISTANCE* mesure la distance qui sépare l'exploitation considérée du marché le plus proche.
- *EXPERT* mesure les heures de discussion entre l'agriculteur et un expert agricole.
- *IRRIG* est une variable binaire qui vaut 1 quand des techniques d'irrigation sont utilisées et 0 dans le cas contraire.
- *PROPR* est une variable binaire qui vaut 1 quand l'agriculteur est propriétaire de l'exploitation et 0 dans le cas contraire.

Pour chaque individu i, le vecteur Z_i est défini de la manière suivante :

$$Z_i = \begin{pmatrix} 1 \\ DETTE_i \\ DISTANCE_i \\ EXPERT_i \\ IRRIG_i \\ PROPR_i \end{pmatrix} = (1 DETTE_i DISTANCE_i EXPERT_i IRRIG_i PROPR_i)'$$

Expliquez pourquoi un agriculteur fait le choix d'utiliser des engrais. Réalisez cet exercice avec TSP.

Solution

Le modèle doit expliquer la probabilité d'utiliser des engrais par les valeurs des variables explicatives observées.

Probit explique cette probabilité de la manière suivante :

$$\begin{aligned} P(ENGRAIS_i = 1) = \Phi(\beta' Z_i) = \Phi(\beta_1 + \beta_2 DETTE_i + \beta_3 DISTANCE_i + \beta_4 EXPERT_i \\ + \beta_5 IRRIG_i + \beta_6 PROPR_i), \quad \text{pour tout } i = 1, 2 \ldots 398 \end{aligned}$$

De son côté, Logit l'explique ainsi :

$$P(ENGRAIS_i = 1) = \Lambda(\beta' Z_i) = \Lambda(\beta_1 + \beta_2 DETTE_i + \beta_3 DISTANCE_i + \beta_4 EXPERT_i + \beta_5 IRRIG_i + \beta_6 PROPR_i), \quad \text{pour tout } i = 1, 2 \ldots 398$$

L'estimation des modèles Probit et Logit se fait au moyen du programme suivant :

```
freq n;
smpl 1 398;
read(file='c:\engrais.xls');
probit engrais c distance dette irrig propr expert;
logit(nchoice=2) engrais c distance dette irrig propr expert;
```

Ce programme fournit les résultats suivants :

```
                              Equation 1
                             ============

                            Probit ESTIMATION

Working space used: 6115

                               STARTING VALUES

                    C     DISTANCE       DETTE        IRRIG
VALUE        0.00000      0.00000      0.00000      0.00000

                PROPR       EXPERT
VALUE        0.00000      0.00000

F=  275.87    FNEW=  217.13    ISQZ=  0 STEP=  1.0000    CRIT=  108.84
F=  217.13    FNEW=  215.27    ISQZ=  0 STEP=  1.0000    CRIT=  3.2874
F=  215.27    FNEW=  214.91    ISQZ=  0 STEP=  1.0000    CRIT= 0.63129
F=  214.91    FNEW=  214.89    ISQZ=  0 STEP=  1.0000    CRIT= 0.33882E-01
F=  214.89    FNEW=  214.89    ISQZ=  0 STEP=  1.0000    CRIT= 0.89120E-04
F=  214.89    FNEW=  214.89    ISQZ=  0 STEP=  1.0000    CRIT= 0.54649E-09

CONVERGENCE ACHIEVED AFTER 6 ITERATIONS

   12 FUNCTION EVALUATIONS.

    Dependent variable: ENGRAIS

  Number of observations = 398      Scaled R-squared = .256259
 Number of positive obs. = 159      LR (zero slopes) = 105.771 [.000]
       Mean of dep. var. = .399497    Schwarz B.I.C. = 232.851
Sum of squared residuals = 71.6068    Log likelihood = -214.892
               R-squared = .250057
Fraction of Correct Predictions =   0.733668

                            Standard
Parameter  Estimate           Error       t-statistic    P-value
C          -1.05029       .141593        -7.41767      [.000]
DISTANCE   -.012237       .012212        -1.00206      [.316]
DETTE      .153616E-03    .726164E-04    2.11545       [.034]
```

```
IRRIG       1.16777      .143337       8.14703      [.000]
PROPR       .308972      .145839       2.11859      [.034]
EXPERT      .016280      .878137E-02   1.85393      [.064]

Standard Errors computed from analytic second derivatives
(Newton)

                                dP/dX
                      0             1
C              0.32091      -0.32091
DISTANCE     0.0037389    -0.0037389
DETTE      -0.000046936   0.000046936
IRRIG         -0.35680       0.35680
PROPR        -0.094403      0.094403
EXPERT      -0.0049742     0.0049742

                                Equation 2
                               ============

                       MULTINOMIAL Logit ESTIMATION

Choice      Frequency   Fraction
0                 239    0.6005  (coefficients normalized to zero)
1                 159    0.3995
Working space used: 6961

                            STARTING VALUES

                 C1      DISTANCE1      DETTE1       IRRIG1
VALUE       0.00000        0.00000     0.00000      0.00000

             PROPR1        EXPERT1
VALUE       0.00000        0.00000

F=  275.87    FNEW=  217.58     ISQZ=  0 STEP=  1.0000    CRIT=  108.84
F=  217.58    FNEW=  215.21     ISQZ=  0 STEP=  1.0000    CRIT=  4.1360
F=  215.21    FNEW=  214.77     ISQZ=  0 STEP=  1.0000    CRIT= 0.75344
F=  214.77    FNEW=  214.73     ISQZ=  0 STEP=  1.0000    CRIT= 0.75332E-01
F=  214.73    FNEW=  214.73     ISQZ=  0 STEP=  1.0000    CRIT= 0.52623E-03
F=  214.73    FNEW=  214.73     ISQZ=  0 STEP=  1.0000    CRIT= 0.21469E-07

CONVERGENCE ACHIEVED AFTER 6 ITERATIONS

  12 FUNCTION EVALUATIONS.

    Dependent variable: ENGRAIS

  Number of observations = 398      Scaled R-squared = .257015
 Number of positive obs. = 159      LR (zero slopes) = 106.095 [.000]
       Mean of dep. var. = .399497    Schwarz B.I.C. = 232.689
Sum of squared residuals = 71.5015    Log likelihood = -214.730
               R-squared = .251146
       Number of Choices =    796
Fraction of Correct Predictions =   0.731156
```

Parameter	Estimate	Standard Error	t-statistic	P-value
C1	-1.75171	.254221	-6.89048	[.000]
DISTANCE1	-.022378	.022533	-.993113	[.321]
DETTE1	.269173E-03	.133790E-03	2.01190	[.044]
IRRIG1	1.92805	.245219	7.86259	[.000]
PROPR1	.528176	.246712	2.14086	[.032]
EXPERT1	.029987	.016441	1.82389	[.068]

Standard Errors computed from analytic second derivatives (Newton)

	dP/dX 0	dP/dX 1
C	0.31601	-0.31601
DISTANCE	0.0040369	-0.0040369
DETTE	-0.000048559	0.000048559
IRRIG	-0.34782	0.34782
PROPR	-0.095283	0.095283
EXPERT	-0.0054096	0.0054096

Pour chacune des méthodes Probit et Logit, on dispose d'un tableau avec une colonne Estimate, qui reprend les coefficients β_i estimés pour chaque variable de la colonne Parameter. Les coefficients estimés par les deux méthodes sont assez proches. Les coefficients de la plupart des variables sont estimés avec une grande précision. Ils sont en effet très significatifs, avec des statistiques *t* élevés et des *p*-values faibles, sauf pour la variable *DISTANCE*. Pour les autres variables, sauf *EXPERT*, on rejette l'hypothèse de nullité du coefficient β_i à un seuil de signification de 5 % puisque leur *p*-value est inférieure à 0,05. Pour EXPERT, la probabilité critique du test Student indique que l'on ne rejette l'hypothèse de nullité du coefficient β_i qu'à un seuil de signification supérieur à 6,8 % dans l'estimation du modèle Logit. Les tableaux dP/dX indiquent, pour chaque variable explicative, dans la colonne de droite (affichant 1 en en-tête), l'impact marginal d'une augmentation de la variable sur la probabilité d'utiliser de l'engrais. Plusieurs observations sont à faire au sujet des déterminants de cette probabilité :

- Elle est d'autant plus faible que la distance entre l'exploitation et le marché est grande : il est inutile d'utiliser des engrais pour produire davantage si l'on ne peut se rendre sur un marché pour y écouler la production sur le marché.
- Elle augmente avec le niveau d'endettement de l'exploitant : un agriculteur très endetté a besoin de produire et de vendre davantage pour pouvoir rembourser.
- Elle augmente avec le temps passé à discuter avec un expert qui explique les avantages à utiliser de l'engrais.
- Elle augmente si l'agriculteur est propriétaire de la parcelle : il est alors davantage impliqué.
- Elle est aussi plus forte quand la parcelle est irriguée : l'irrigation favorise l'utilité de l'engrais.

La statistique *R*-squared correspond au pseudo-R^2. La statistique *LR* (zero slopes) est le test statistique Chi-2.

Exercice 2 Prendre la décision d'utiliser des engrais avec Easyreg

Énoncé

Estimez à nouveau les modèles de choix d'utilisation d'engrais avec le logiciel Easyreg.

Solution

Lecture des données

Ouvrez le fichier ENGRAIS.xls avec EXCEL, cliquez sur Fichier puis sur Enregistrer sous, et sélectionnez CSV (séparateur point-virgule) pour le type de fichier. Laissez Disque local C:\ dans la zone Enregistrer dans et conservez UKL comme nom de fichier. Cliquez sur Enregistrer. Le logiciel fabrique alors un fichier ENGRAIS.csv dans C:\. Ouvrez Easyreg et cliquez successivement sur File/Choose an input file/Choose an Excel file in CSV format. Dans la fenêtre Guided tour qui apparaît, cliquez sur la réponse de votre choix ; si vous demandez le Guided tour, il faut ensuite réduire la fenêtre Guided tour on importing Excel files in CSV format pour pouvoir continuer à travailler avec Easyreg (et voir la fenêtre About CSV files). Dans la fenêtre About CSV files, cliquez sur Continue. Dans la fenêtre Find a CSV file qui s'affiche, sélectionnez Manually dans Search options, C: dans Choose drive, et C:\ dans Choose drive/directory. Cochez Double-click to choose file and continue. Double-cliquez sur C:\ engrais.csv dans Files found. La fenêtre Data type of the CSV file apparaît. Cochez Annual time series, même si les données sont journalières. Saisissez 1 dans Year of first observation et cliquez sur OK. Cliquez sur Confirm. Cliquez sur Continue. L'écran Analyzing the CSV file s'affiche. Cliquez sur Code OK. La fenêtre Data written to files apparaît. Cliquez sur Continue. La fenêtre Check the data lui succède. Cliquez sur Continue. Sur About the start-up folder EASYREG.DAT, cliquez sur Continue. Choose drive and folder to start-up EasyReg apparaît. Sélectionnez C:\. Ensuite saisissez par exemple ENGRAIS dans Make a new sub-folder in the Current folder – Enter the folder name, et cliquez sur OK. Cliquez sur Start Easyreg in current folder.

Estimation des modèles

Cliquez successivement sur Menu/Single equation models/Discrete dependent variable. Dans la fenêtre Guided tour qui apparaît, cliquez sur la réponse de votre choix ; si vous demandez le Guided tour, il faut ensuite réduire la fenêtre Guided tour on discrete dependent pour pouvoir continuer à travailler avec Easyreg. Dans la fenêtre Select all variables in your limited, double-cliquez sur les variables *EXPERT*, *DISTANCE*, *IRRIG*, *DETTE*, *PROPR* et *ENGRAIS*. Ensuite, cliquez sur Selection OK, puis sur NO en réponse à la question Do you want to use a subset of observations, et sur Continue. Dans la fenêtre Select the dependent (Y) variable in your limited dependent, double-cliquez sur *ENGRAIS*. Ensuite cliquez deux fois sur Continue.

Dans la fenêtre Select the independent (X) variables, double-cliquez sur *EXPERT*, *DISTANCE*, *IRRIG*, *DETTE*, *PROPR*. Cliquez sur Selection OK. Dans la fenêtre Select laggeg dependent Y variables as X variables, cliquez sur Skip all. Dans la fenêtre Select (lagged) independent variables, double-cliquez sur *EXPERT*, puis cliquez sur Selection OK. Dans la fenêtre Select (lagged) independent variables, double-cliquez sur *DISTANCE*, puis cliquez sur Selection OK. Dans la fenêtre Select (lagged) independent variables, double-cliquez sur *IRRIG*, puis cliquez sur Selection OK.

Dans la fenêtre Select (lagged) independent variables, double-cliquez sur *DETTE*, puis cliquez sur Selection OK. Dans la fenêtre Select (lagged) independent variables, double-cliquez sur *PROPR*, puis cliquez sur Selection OK. Dans la fenêtre Select deterministic X

variables, cliquez sur Continue. Dans la fenêtre Model variables, cliquez de nouveau sur Continue. Dans la fenêtre Discrete dependent variable, cliquez sur Logit. Vous obtenez les résultats suivants :

```
Logit model:
Dependent variable:
Y = ENGRAIS

Characteristics:
ENGRAIS
  First observation = 1
  Last observation  = 398
  Number of usable observations: 398
  Minimum value: 0.0000000E+000
  Maximum value: 1.0000000E+000
  Sample mean:   3.9949749E-001
  This variable is a zero-one dummy variable.
  A Probit or Logit model is suitable.

X variables:
X(1) = EXPERT
X(2) = DISTANCE
X(3) = IRRIG
X(4) = DETTE
X(5) = PROPR
X(6) = 1

Frequency of y = 1:39.95 %
Frequency of y = 0:60.05 %
Model: P(Y=1|x) = F(b(1)x(1)+..+b(6)x(6))
Chosen option: F(u) = 1/[1+EXP(-u)] (Logit model)
Newton iteration succesfully completed after 6 iterations
Last absolute parameter change = 0.0000
Last percentage change of the likelihood = 0.0000

Maximum likelihood estimation results:
Variable                       ML estimate of b(.) (t-value)
x(1)=EXPERT                    b(1)=   0.0299873 (1.82)
                                     [p-value = 0.06817]
x(2)=DISTANCE                  b(2)=  -0.0223412 (-0.99)
                                     [p-value = 0.32127]
x(3)=IRRIG                     b(3)=   1.9280423 (7.86)
                                     [p-value = 0.00000]
x(4)=DETTE                     b(4)=   0.0002692 (2.01)
                                     [p-value = 0.04424]
x(5)=PROPR                     b(5)=   0.5281730 (2.14)
                                     [p-value = 0.03229]
x(6)=1                         b(6)=  -1.7518778 (-6.89)
                                    [p-value = 0.00000]
[The two-sided p-values are based on the normal approximation]

Log likelihood:      -2.14730755414E+002
Sample size (n):                     398

If the model is correctly specified then the maximum likelihood
parameter estimators b(1),.., b(6), minus their true values, times the
square root of the sample size n, are (asymptotically) jointly normally
distributed with zero mean vector and variance matrix :
```




```
 1.07591140E-01 8.01242020E-03 -1.35148898E-01 -9.13439369E-05 -1.17571981E-01
-1.48712326E-01
 8.01242020E-03 2.01931716E-01 1.52168797E-01 -1.22568933E-04 -1.80605555E-01
-1.02391613E+00
 -1.35148898E-01 1.52168797E-01 2.39329261E+01 -9.19966801E-04 2.90660897E+00
-1.55812324E+01
 -9.13439369E-05 -1.22568933E-04 -9.19966801E-04 7.12404487E-06 1.88806501E-03
-2.17737458E-03
 -1.17571981E-01 -1.80605555E-01 2.90660897E+00 1.88806501E-03 2.42249161E+01
-1.09274144E+01
 -1.48712326E-01 -1.02391613E+00 -1.55812324E+01 -2.17737458E-03 -1.09274144E+01
2.57228252E+01
```

Les valeurs estimées des coefficients β_i des différentes variables sont données par les valeurs des ML estimate of b(). Les coefficients estimés sont les mêmes que ceux obtenus avec TSP.

Cliquez sur Done. Refaites toute la procédure mais en choisissant cette fois Probit au lieu de Logit dans la fenêtre Discrete dependent variable. Vous obtenez les résultats suivants :

```
Probit model:
Dependent variable:
Y = ENGRAIS

Characteristics:
ENGRAIS
  First observation = 1
  Last observation = 398
  Number of usable observations: 398
  Minimum value: 0.0000000E+000
  Maximum value: 1.0000000E+000
  Sample mean:   3.9949749E-001
  This variable is a zero-one dummy variable.
  A Probit or Logit model is suitable.

X variables:
X(1) = EXPERT
X(2) = DISTANCE
X(3) = IRRIG
X(4) = DETTE
X(5) = PROPR
X(6) = 1

Frequency of y = 1:39.95 %
Frequency of y = 0:60.05 %
Model: P(Y=1|x) = F(b(1)x(1)+..+b(6)x(6))
Chosen option: F(u) = c.d.f. of N(0,1) distr. (Probit model)
Newton iteration succesfully completed after 6 iterations
Last absolute parameter change = 0.0000
Last percentage change of the likelihood = 0.0000

Maximum likelihood estimation results:
Variable                       ML estimate of b(.) (t-value)
x(1)=EXPERT                    b(1)=   0.0162802 (1.85)
                                     [p-value = 0.06376]
x(2)=DISTANCE                  b(2)=  -0.0122180 (-1.00)
                                     [p-value = 0.31693]
x(3)=IRRIG                     b(3)=   1.1677607 (8.15)
                                     [p-value = 0.00000]
```

```
x(4)=DETTE                      b(4)=   0.0001536 (2.12)
                                      [p-value = 0.03441]
x(5)=PROPR                      b(5)=   0.3089698 (2.12)
                                      [p-value = 0.03413]
x(6)=1                          b(6)=  -1.0503782 (-7.42)
                                      [p-value = 0.00000]
[The two-sided p-values are based on the normal approximation]

Log likelihood:      -2.14893153349E+002
Sample size (n):                      398

If the model is correctly specified then the maximum likelihood
parameter estimators b(1),.., b(6), minus their true values, times the
square root of the sample size n, are (asymptotically) jointly normally
distributed with zero mean vector and variance matrix:

 3.06929185E-02 2.19694628E-03 -6.18845251E-02 -2.88036261E-05 -5.07177648E-02
-2.78266609E-02
 2.19694628E-03 5.93194461E-02 6.22160725E-02 -3.23305403E-05  -6.51424731E-02
-3.11861222E-01
-6.18845251E-02 6.22160725E-02 8.17723437E+00 -3.90437779E-04 6.54847927E-01
-4.79552208E+00
-2.88036261E-05 -3.23305403E-05 -3.90437779E-04 2.09877359E-06 6.65710191E-04
-6.76957123E-04
-5.07177648E-02 -6.51424731E-02 6.54847927E-01 6.65710191E-04 8.46501714E+00
-3.51616503E+00
-2.78266609E-02 -3.11861222E-01 -4.79552208E+00 -6.76957123E-04 -3.51616503E+00
7.97943435E+00
```

EXERCICE 3 MODÉLISATION DU TAUX DE CHANGE ENTRE LA LIVRE STERLING ET LE DOLLAR

Énoncé

Le fichier UKL.XLS réunit 3 897 observations quotidiennes successives du taux de change entre la livre sterling et le dollar. Il s'agit d'une feuille de calcul Excel créée de manière que TSP puisse la lire. Elle est placée dans le dossier C:\. Dans ce fichier, la série de taux de change est appelée UKL. Vérifiez si le marché des changes s'est avéré efficient pour ce taux de change, et si celui-ci montre une volatilité conditionnelle autorégressive. Travaillez avec TSP.

Solution

Vérifier l'efficience du marché

Si le marché a l'efficience d'utiliser à tout moment l'ensemble de l'information disponible pour former le taux de change, la variation de ce dernier d'une période à l'autre ne devrait résulter que d'un flux d'information nouvelle, imprévisible durant les périodes précédentes. La variation du taux de change ne devrait donc pas être liée à ses valeurs passées ; elle ne devrait pas être autocorrélée.

Il est intéressant de vérifier cela en regardant l'autocorrélogramme empirique simple de la variation du taux de change. Pour cela, on fait appel au programme TSP suivant :

```
freq n;
smpl 1 3898;
```

```
read(file='c:\ukl.xls');
smpl 2 3898;
dukl=ukl-ukl(-1);
bjident(nlag=12) dukl;
```

On génère la variable `dukl`, qui est la variation du taux de change `ukl`. La commande `bjident` permet d'obtenir l'autocorrélogramme simple pour cette série dexch. Les résultats sont les suivants :

```
======================
Box-Jenkins procedures
Procedure   BJIDENT
======================

                           OPTIONS FOR THIS ROUTINE
                           ========================

IAC      = FALSE         NDIFF   = 0            NLAG = 12
NLAGP    = 10            NSDIFF  = 0            NSPAN = 0
PACMETH  = BURG          PLOT    = TRUE         PLOTAC = TRUE
PLTRAW   = FALSE         SILENT  = FALSE

                                Autocorrelations
                                ================

Series: DUKL
        Mean = -0.74416330E-06
  Std. Error =  0.37493810E-02

                  Lags
Autocorrelations        -0.215E-01 0.108E-01-0.128E-01 0.288E-01 0.259E-01
Standard Errors   1- 5   0.160E-01 0.160E-01 0.160E-01 0.160E-01 0.160E-01
Q-statistics              1.80      2.25      2.89      6.14      8.75
Autocorrelations        -0.341E-01-0.129E-01-0.482E-02 0.141E-01 0.351E-01
Standard Errors   6-10   0.161E-01 0.161E-01 0.161E-01 0.161E-01 0.161E-01
Q-statistics              13.3     13.9      14.0      14.8      19.6
Autocorrelations        -0.217E-01-0.173E-01
Standard Errors   11-12  0.161E-01 0.161E-01
Q-statistics              21.5     22.6

                              Partial Autocorrelations
                              ========================

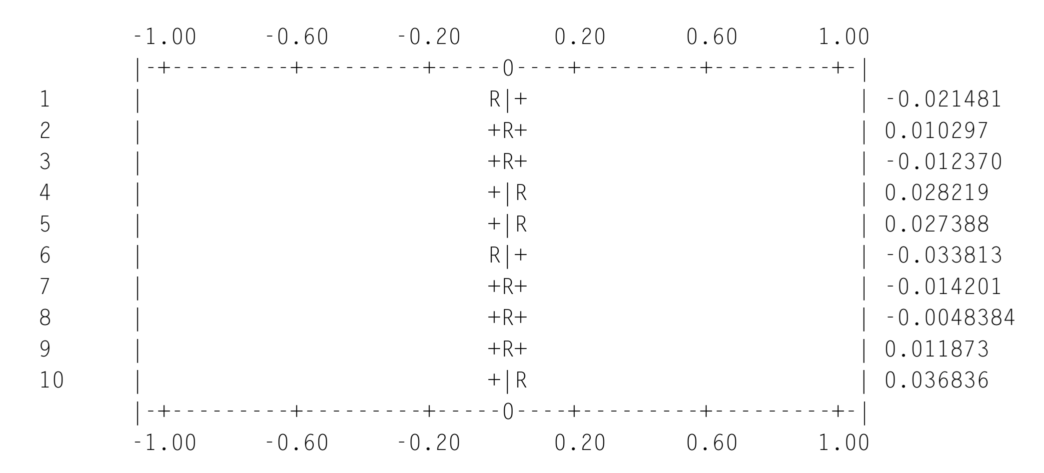
Series: DUKL
  Standard Error of Autocorrelations =  0.16018978E-01

                   Lags
Partial Autocorrs  1- 5 -0.215E-01 0.103E-01-0.124E-01 0.282E-01 0.274E-01
Partial Autocorrs  6-10 -0.338E-01-0.142E-01-0.484E-02 0.119E-01 0.368E-01

Autocorrelation Function of:  DUKL
```

```
       -1.00     -0.60     -0.20       0.20      0.60      1.00
       |-+---------+---------+-----0-----+---------+---------+-|
1      |                          R|+                           | -0.021479
2      |                          +R+                           | 0.010753
3      |                          +R+                           | -0.012811
4      |                          +|R                           | 0.028838
5      |                          +|R                           | 0.025880
6      |                          R|+                           | -0.034112
7      |                          +R+                           | -0.012878
8      |                          +R+                           | -0.0048205
9      |                          +R+                           | 0.014079
10     |                          +|R                           | 0.035085
11     |                          R|+                           | -0.021683
12     |                          +R+                           | -0.017293
       |-+---------+---------+-----0-----+---------+---------+-|
       -1.00     -0.60     -0.20       0.20      0.60      1.00
```

```
Partial Autocorrelation Function of:  DUKL
```



```
       -1.00     -0.60     -0.20       0.20      0.60      1.00
       |-+---------+---------+-----0-----+---------+---------+-|
1      |                          R|+                           | -0.021481
2      |                          +R+                           | 0.010297
3      |                          +R+                           | -0.012370
4      |                          +|R                           | 0.028219
5      |                          +|R                           | 0.027388
6      |                          R|+                           | -0.033813
7      |                          +R+                           | -0.014201
8      |                          +R+                           | -0.0048384
9      |                          +R+                           | 0.011873
10     |                          +|R                           | 0.036836
       |-+---------+---------+-----0-----+---------+---------+-|
       -1.00     -0.60     -0.20       0.20      0.60      1.00
```

La variation `dukl` de `ukl` n'est pas autocorrélée ; cela confirme l'efficience du marché des changes. Graphiquement, les coefficients d'autocorrélation empiriques sont à l'intérieur de l'intervalle dans lequel on ne rejette pas l'hypothèse que les coefficients d'autocorrélation de population sont nuls. Les tests Q sont inférieurs à leurs valeurs critiques respectives à 5 %.

Vérifier le caractère autorégressif de la volatilité

Le fait que le niveau de la variation du taux de change ne dépende pas des niveaux précédents n'empêche pas la variance du taux de change de dépendre des variances précédentes. Autrement dit, même avec un niveau de taux de change non autocorrélé, la variance du taux de change peut être autocorrélée. La variance $V(X_t)$ d'une variable X_t est égale à $E\big((X_t - E(X_t)\big)2$. Dans cette application, X_t est égal à $DUKL_t$, dont la moyenne est presque nulle, ce qui suggère que $E(DUKL_t) \sim 0$. Dès lors, $V(DUKL_t) \sim E(DUKL_t^2)$. Vérifier que $V(DUKL_t)$ est autocorrélé revient à vérifier que $E(DUKL_t^2)$ est autocorrélé. On n'observe pas $E(DUKL_t^2)$, mais bien $DUKL_t^2$ à chaque période. Il suffit donc de vérifier que $DUKL_t^2$ est autocorrélé en regardant l'autocorrélogramme empirique simple du carré de la variation du taux de change.

Pour cela, on utilise le programme TSP suivant :

```
freq n;
smpl 1 3898;
read(file='c:\ukl.xls');
smpl 2 3898;
dukl=ukl-ukl(-1);
dukl2=dukl**2;
bjident(nlag=12) dukl2;
```

Les résultats sont :

```
======================
Box-Jenkins procedures
Procedure   BJIDENT
======================

                         OPTIONS FOR THIS ROUTINE
                         ========================

IAC     = FALSE          NDIFF    = 0             NLAG = 12
NLAGP   = 10             NSDIFF   = 0            NSPAN = 0
PLOT    = TRUE           PLOTAC   = TRUE        PLTRAW = FALSE

                              Autocorrelations
                              ================

                                            0     0 0

                                Series (1-B) (1-B )   DEXCH2
                                              Mean =  0.75927506E-03
                                    Standard Error =  0.15345252E-02

                  Lags

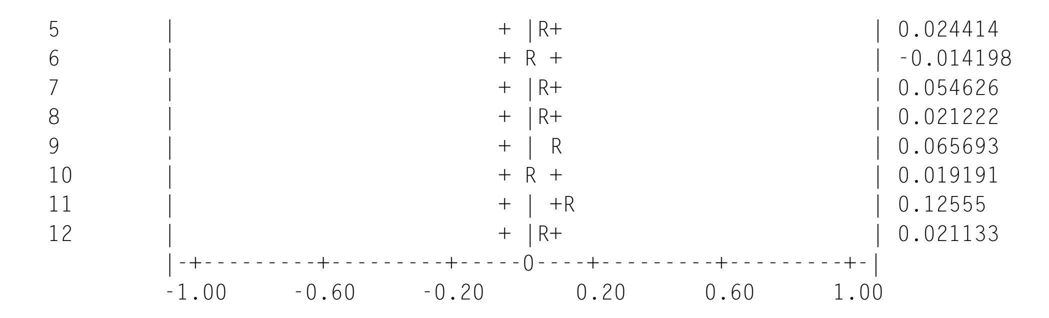
Autocorrelations         0.732E-01 0.528E-01 0.151     0.129      0.244E-01
Standard Errors   1- 5   0.462E-01 0.464E-01 0.466E-01 0.476E-01  0.483E-01
Q-statistics              2.53      3.85     14.7      22.6       22.8
Autocorrelations        -0.142E-01 0.546E-01 0.212E-01 0.657E-01 0.192E-01
Standard Errors   6-10   0.483E-01 0.484E-01 0.485E-01 0.485E-01 0.487E-01
Q-statistics             22.9      24.4      24.6      26.6      26.8
Autocorrelations         0.126     0.211E-01
Standard Errors  11-12   0.487E-01 0.494E-01
Q-statistics             34.4      34.6

0     0 0

            Autocorrelation Function of the series (1-B) (1-B )   DEXCH2

          -1.00     -0.60     -0.20       0.20      0.60      1.00
          |-+---------+---------+-----0----+---------+---------+-|
 1        |                         + | R                        | 0.073215
 2        |                         + |R+                        | 0.052828
 3        |                         + | + R                      | 0.15116
 4        |                         + | +R                       | 0.12874
```

Exercices


```
5        |                         + |R+                        | 0.024414
6        |                         + R +                        | -0.014198
7        |                         + |R+                        | 0.054626
8        |                         + |R+                        | 0.021222
9        |                         + |  R                       | 0.065693
10       |                         + R +                        | 0.019191
11       |                         + | +R                       | 0.12555
12       |                         + |R+                        | 0.021133
         |-+---------+---------+-----0----+---------+---------+-|
         -1.00     -0.60     -0.20      0.20      0.60      1.00
```

L'autocorrélation de `dukl2` suggère que `dukl2` est autocorrélé et donc que `dukl` est un processus de type *ARCH* ou *GARCH*. On vérifie plus rigoureusement l'existence d'un tel effet *ARCH* dans le processus de *DEXCH* au moyen d'un test formel. Pour cela, on exécute le programme suivant :

```
freq n;
smpl 1 3898;
read(file='ukl.xls');
smpl 2 3898;
dukl=ukl-ukl(-1);
dukl2=dukl**2;
smpl 12 3898;
noplots;
olsq dukl2 c dukl2(-1)- dukl2(-10);
const lm;
set lm=@rsq*@nob;
print lm;
```

L'instruction `olsq` réalise l'estimation par MCO d'un modèle linéaire où la variable dépendante est `dukl2` et les variables explicatives sont à peu près les dix premières valeurs retardées de `dukl2` et une constante. L'instruction suivante définit un paramètre `lm` comme une constante. L'instruction `set` affecte la valeur de cette constante. TSP utilise la notation `@rsq` pour appeler le R^2 des résidus de la régression précédente estimée par MCO. Le terme `@nob` correspond à l'effectif des observations sur lesquelles cette régression a été estimée. Les résultats sont les suivants :

```
                              Equation 1
                             ============

                Method of estimation = Ordinary Least Squares

Dependent variable:  DUKL2
Current sample:  12 to 3898
Number of observations:  3887

       Mean of dep. var. = .140807E-04
  Std. dev. of dep. var. = .340999E-04
Sum of squared residuals = .431192E-05
   Variance of residuals = .111247E-08
```

```
Std. error of regression = .333536E-04
               R-squared = .045752
      Adjusted R-squared = .043290
            LM het. test = 66.3115 [.000]
           Durbin-Watson = 2.00515 [<.663]
         Durbin's h alt. = -4.02320 [.000]
        Jarque-Bera test = .105864E+07 [.000]
         Ramsey's RESET2 = 16.4058 [.000]
         F (zero slopes) = 18.5839 [.000]
          Schwarz B.I.C. = -34513.2
          Log likelihood = 34558.6

             Estimated    Standard
Variable    Coefficient     Error       t-statistic   P-value
C           .758102E-05   .747847E-06   10.1371       [.000]
DUKL2(-1)   .080748       .016049       5.03122       [.000]
DUKL2(-2)   .069234       .016052       4.31301       [.000]
DUKL2(-3)   .028115       .016091       1.74725       [.081]
DUKL2(-4)   .048134       .016097       2.99018       [.003]
DUKL2(-5)   .033812       .016056       2.10591       [.035]
DUKL2(-6)   .086681       .016056       5.39883       [.000]
DUKL2(-7)   -.140970E-02  .016097       -.087573      [.930]
DUKL2(-8)   -.216749E-02  .016091       -.134702      [.893]
DUKL2(-9)   .078521       .016052       4.89150       [.000]
DUKL2(-10)  .039911       .016050       2.48671       [.013]

LM = 177.83953
```

Pour tester l'hypothèse d'absence d'effet *ARCH*, il suffit de mettre la valeur du test *LM* en regard de la valeur critique d'une loi Chi-2 à 10 degrés de liberté. Selon cette hypothèse, `dukl2` ne dépend pas de ses valeurs passées. Cela implique que la volatilité n'est pas autorégressive. Ici, le test vaut 177.83953, soit un résultat supérieur à la valeur critique 18.3070. On rejette donc l'hypothèse que tous les coefficients des `dukl2(-i)` sont nuls : on est bien en présence d'un effet *ARCH*.

On peut vérifier que `dukl2` suit un processus *ARMA* en estimant un modèle *ARMA*(1, 1) pour `dukl2` :

```
freq n;
smpl 1 3898;
read(file='c:\ukl.xls');
smpl 2 3898;
dukl=ukl-ukl(-1);
dukl2=dukl**2;
bjest(nar=1,nma=1,constant,noplot,nocumplo) dukl2 start phi(1) .921 theta(1) .871;
```

Ce programme donne les résultats suivants :

```
======================
Box-Jenkins procedures
Procedure   BJEST
======================
```

```
Working space used: 47367

                                       STARTING VALUES

                  PHI1        THETA1        CONSTANT
VALUE          0.92100       0.87100     1.30702D-06

F=  -34639.     FNEW=  -34650.     ISQZ=  1 STEP= 0.25000     CRIT=  39.047
F=  -34650.     FNEW=  -34652.     ISQZ=  0 STEP=  1.0000     CRIT=  22.165
F=  -34652.     FNEW=  -34663.     ISQZ=  0 STEP=  1.0000     CRIT=  21.513
F=  -34663.     FNEW=  -34663.     ISQZ=  0 STEP=  1.0000     CRIT= 0.28527
F=  -34663.     FNEW=  -34663.     ISQZ=  0 STEP=  1.0000     CRIT= 0.90534E-02
F=  -34663.     FNEW=  -34663.     ISQZ=  0 STEP=  1.0000     CRIT= 0.60963E-03

CONVERGENCE ACHIEVED AFTER 6 ITERATIONS

  87 FUNCTION EVALUATIONS.

                         Results of Box-Jenkins Estimation
                         =================================

Dependent variable: DUKL2
Current sample:  2 to 3898
Number of observations:  3897

       Mean of dep. var. = .140543E-04  Adjusted R-squared = .050596
  Std. dev. of dep. var. = .340606E-04        LM het. test = 42.0314 [.000]
Sum of squared residuals = .428896E-05       Durbin-Watson = 1.94763
   Variance of residuals = .110143E-08      Schwarz B.I.C. = -34650.5
Std. error of regression = .331878E-04      Log likelihood = 34662.9
               R-squared = .051084

                         Standard
Parameter  Estimate        Error       t-statistic    P-value
PHI1       .986099       .454929E-02   216.759        [.000]
THETA1     .947381       .873916E-02   108.406        [.000]
CONSTANT   .193660E-06   .696202E-07   2.78166        [.005]

Standard Errors computed from  quadratic form of analytic first
derivatives (Gauss)

                       Autocorrelations of the Residuals

                     1             2             3            4
Autocorr      0.026178      0.016287     -0.016528    0.0036132
Q-stat        2.67265       3.70752        4.77345      4.82441
P-value        .             .            0.028902     0.089618

                      5             6             7            8
Autocorr      -0.010011      0.040464     -0.040680    -0.042808
Q-stat          5.21568      11.60957      18.07353     25.23314
P-value         0.15667      0.020503     0.0028557   0.00030913

                      9            10            11           12
Autocorr       0.029808    0.00013981      0.027821     0.014415
Q-stat         28.70552      28.70560      31.73197     32.54462
P-value      0.00016371    0.00035711    0.00022152   0.00032476
```

Il faut donc utiliser l'instruction ARCH pour estimer un modèle GARCH. La valeur de `nar` correspond à la valeur de *p*, et celle de `nma` à celle de *q* :

```
                          Equation 2
                         ============

                       GARCH ESTIMATION


                    OPTIONS FOR THIS ROUTINE
                    ========================

  E2INIT  = HINIT       GT      =            HEXP     = 0.50000
  HINIT   = SSR         MEAN    = FALSE      NAR      = 1
  NMA     = 1           RELAX   = FALSE      UNCOND   = FALSE
  ZERO    = TRUE
Working space used: 54925

                              STARTING VALUES

                     C         ALPHA0       ALPHA1       BETA1
VALUE     -7.44163D-07    0.000014054      0.00000     0.00000

F=  -16240.    FNEW=  -16261.    ISQZ=  0 STEP=  1.0000    CRIT=  30.959
F=  -16261.    FNEW=  -16391.    ISQZ=  0 STEP=  1.0000    CRIT=  71.971
F=  -16391.    FNEW=  -16440.    ISQZ=  2 STEP= 0.25000    CRIT=  195.69
F=  -16440.    FNEW=  -16463.    ISQZ=  3 STEP= 0.12500    CRIT=  301.39
F=  -16463.    FNEW=  -16471.    ISQZ=  0 STEP=  1.0000    CRIT=  21.483
F=  -16471.    FNEW=  -16473.    ISQZ=  0 STEP=  1.0000    CRIT=  2.6573
F=  -16473.    FNEW=  -16473.    ISQZ=  0 STEP=  1.0000    CRIT= 0.15301
F=  -16473.    FNEW=  -16473.    ISQZ=  0 STEP=  1.0000    CRIT= 0.89389E-03
F=  -16473.    FNEW=  -16473.    ISQZ=  0 STEP=  1.0000    CRIT= 0.47975E-07

CONVERGENCE ACHIEVED AFTER 9 ITERATIONS

   23 FUNCTION EVALUATIONS.

Dependent variable: DUKL
Current sample:  2 to 3898
Number of observations:  3897

(Statistics based on transformed data)
       Mean of dep. var. = -.458069E-04
  Std. dev. of dep. var. = .337224E-02
Sum of squared residuals = .044307
   Variance of residuals = .113724E-04
Std. error of regression = .337230E-02
               R-squared = .477087E-02
      Adjusted R-squared = .477087E-02
           Durbin-Watson = 2.02142 [<1.00]
        Jarque-Bera test = 861.453 [.000]

(Statistics based on original data)
       Mean of dep. var. = -.744163E-06
  Std. dev. of dep. var. = .374938E-02
Sum of squared residuals = .054786
```

```
   Variance of residuals = .140622E-04
Std. error of regression = .374995E-02
               R-squared = .954909E-03
      Adjusted R-squared = .954909E-03
           Durbin-Watson = 2.04213
          Schwarz B.I.C. = -16456.5
          Log likelihood = 16473.0
            Number of observations in LogL =   3897
              Initial observations dropped =      0
              Est. initial values for H(t) =      0
                   Initial values for H(t) =  0.14059E-04

                          Standard
Parameter   Estimate       Error        t-statistic   P-value
C           -.662670E-04  .550785E-04   -1.20314      [.229]
ALPHA0      .163120E-06   .733031E-07   2.22528       [.026]
ALPHA1      .040517       .962832E-02   4.20813       [.000]
BETA1       .947803       .013215       71.7222       [.000]

Standard Errors computed from analytic first and second derivatives
(Eicker-White)
```

Exercice 4 Modélisation du taux de change avec Easyreg

Énoncé

Estimez un modèle *GARCH*(1, 1) pour la variation du taux de change entre la livre sterling et le dollar, en utilisant Easyreg et les données du fichier UKL.XLS précédent.

Solution

Voici comment procéder :

Lecture des données

Ouvrez le fichier UKL.XLS sous Excel, cliquez sur Fichier puis sur Enregistrer sous, et sélectionnez CSV (séparateur point-virgule) comme type de fichier. Laissez Disque local C:\ dans la zone Enregistrer dans et conservez UKL comme nom du fichier. Cliquez sur Enregistrer. Le logiciel crée un fichier UKL.csv dans C:\. Ouvrez Easyreg et cliquez successivement sur File/Choose an input file/Choose an Excel file in CSV format.

Dans la fenêtre Guided tour qui apparaît, cliquez sur la réponse de votre choix ; si vous demandez le Guided tour, il faut ensuite réduire la fenêtre Guided tour on importing Excel files in CSV format pour pouvoir continuer à travailler avec Easyreg (et voir la fenêtre About CSV files). Dans la fenêtre About CSV files, cliquez sur Continue. La fenêtre Find a CSV file s'ouvre : sélectionnez Manually dans Search options, C: dans Choose drive, et C:\ dans Choose drive/directory. Cochez Double-click to choose file and continue. Double-cliquez sur C:\ UKL.csv dans Files found. Data type of the CSV file s'affiche. Cochez Annual time series, même si les données sont journalières. Saisissez 1 dans Year of first observation et cliquez sur OK. Ensuite cliquez sur Confirm, puis sur Continue. Analyzing the CSV file s'affiche. Cliquez sur Code OK.

Lorsque les fenêtres suivantes apparaissent, cliquez sur Continue : Data written to files ; Check the data ; About the start-up folder EASYREG.DAT. Maintenant, Choose drive and

folder to start-up EasyReg apparaît. Sélectionnez C:\. Ensuite saisissez EXCHANGEL dans Make a new sub-folder in the Current folder – Enter the folder name et cliquez sur OK. Cliquez sur Start Easyreg in current folder.

Transformation des données

Cliquez successivement sur Menu/Input/Transform variables. L'écran Transformation of variables s'affiche. Cliquez sur Time series transformations puis sur Difference : $x(t) - x(t - m)$. Le logiciel affiche $m = 1$. Cliquez de nouveau sur Difference : $x(t) - x(t - m)$, sur Continue et sur OK. Enfin, cliquez sur Done. Le logiciel a créé une nouvelle variable $DIF1[UKL]$.

Estimation du modèle

Cliquez sur Menu, sur Single equation models puis sur Linear regression models. Dans la fenêtre Guided tour qui apparaît, cliquez sur la réponse de votre choix ; si vous demandez le Guided tour, il faut ensuite réduire la fenêtre Guided tour on ordinary least squeres estimation and extensions pour pouvoir continuer à travailler avec Easyreg. Dans la fenêtre Select all variables in your linear regression model, double-cliquez sur DIF1[UKL] puis cliquez sur Selection OK.

Dans la fenêtre Selected variables for OLS estimation, cliquez sur NO en réponse à la question « Do you want to use a subset of observations ». Ensuite cliquez sur Continue et de nouveau sur Continue. Dans la fenêtre Select the independent (X) variables, ne sélectionnez aucune variable et cliquez simplement sur Selection OK. Dans la fenêtre Select laggeg dependent Y variables as X variables, cliquez sur Skip all. Dans la fenêtre Select deterministic X variables, cliquez sur Continue. Dans la fenêtre suivante, le logiciel affiche $q = 0$. Cliquez sur q OK puis sur Continue. Dans la fenêtre Model variables, cliquez sur Continue. Les résultats de la régression de la variation du taux de change sur une constante apparaissent :

```
Dependent variable:
Y = DIF1[UKL]

Characteristics:
DIF1[UKL]
  First observation = 2
  Last observation  = 3898
  Number of usable observations: 3897
  Minimum value: -2.5300000E-002
  Maximum value: 2.3500000E-002
  Sample mean: -7.4416218E-007

X variables:
X(1) = 1

Model:
Y = b(1)X(1) + U,
where U is the error term, satisfying
E[U|X(1)] = 0.
```

```
OLS estimation results
Parameters  Estimate    t-value   H.C. t-value(*)
                       [p-value]   [H.C. p-value]
b(1)         0.00000      -0.012           -0.012
                       [0.99011]       [0.99011]

(*) Based on White's heteroskedasticity consistent variance matrix.
[The two-sided p-values are based on the normal approximation]
Effective sample size (n) = 3897
Variance of the residuals = 0.000014
Standard error of the residuals = 0.003749
Residual sum of squares (RSS)= 0.054769
Total sum of squares (TSS) = 0.054769
R-square = 0.000000
Adjusted R-square = 0.000000

The Durbin-Watson test requires an intercept
and at least one other X variable, and is
therefore not reported.

Jarque-Bera/Salmon-Kiefer test = 2440.716711
Null hypothesis: The errors are normally distributed
Null distribution: Chi-square(2))
p-value = 0.00000
Significance levels:         10 %      5 %
Critical values:            4.61    5.99
Conclusions:              reject  reject

The Breusch-Pagan test requires an intercept
and at least one other X variable, and is
therefore not reported.

Information criteria:
Akaike:             -1.11718E+01
Hannan-Quinn:       -1.11712E+01
Schwarz:            -1.11681E+01
```

Cliquez sur Continue. Dans l'écran Model type : OLS : what to do next?, cliquez sur Options et puis sur Re-estimate the model with GARCH errors. Dans l'écran GARCH error structure specification, cliquez sur $r(1, 1)$ et sur $r(2, 1)$. La valeur maximale de i reprise pour les coefficients $r(1, i)$ est égale à q et la valeur maximale de l reprise pour les coefficients $r(2, l)$ est égale à p. En repartant de $r(1, 1)$ et $r(2, 1)$, fixez $q = 1$ et $p = 1$. Dans l'écran Simplex iteration, cliquez sur start SIMPLEX iteration. Le logiciel affiche les valeurs estimées des coefficients dans Parameter Values. Vous pouvez cliquer ensuite sur restart SIMPLEX iteration. Le logiciel affiche alors les nouvelles valeurs estimées des coefficients. Faites cela plusieurs fois successivement jusqu'à ce que les valeurs des paramètres ne changent plus. Cliquez sur Done with SIMPLEX iteration. Ensuite cliquez sur Continue.

Les résultats de l'estimation s'affichent :

```
Dependent variable:
Y = DIF1[UKL]

Characteristics:
DIF1[UKL]
  First observation = 2
  Last observation  = 3898
  Number of usable observations: 3897
  Minimum value: -2.5300000E-002
  Maximum value: 2.3500000E-002
  Sample mean:   -7.4416218E-007

X variables:
X(1) = 1

Model:
Y = b(1)X(1) + U,
where U is the error term, satisfying
E[U|X(1)] = 0.

OLS estimation results
Parameters  Estimate     t-value    H.C. t-value(*)
                                    [p-value]   [H.C.  p-value]
b(1)             0.00000      -0.012        -0.012
                                    [0.99011]         [0.99011]

(*) Based on White's heteroskedasticity consistent variance matrix.
[The two-sided p-values are based on the normal approximation]
Effective sample size (n) = 3897
Variance of the residuals = 0.000014
Standard error of the residuals = 0.003749
Residual sum of squares (RSS)= 0.054769
Total sum of squares (TSS) = 0.054769
R-square = 0.000000
Adjusted R-square = 0.000000

The Durbin-Watson test requires an intercept
and at least one other X variable, and is
therefore not reported.

Jarque-Bera/Salmon-Kiefer test = 2440.716711
Null hypothesis: The errors are normally distributed
Null distribution: Chi-square(2))
p-value = 0.00000
Significance levels:          10 %      5 %
Critical values:             4.61    5.99
Conclusions:               reject  reject

The Breusch-Pagan test requires an intercept
and at least one other X variable, and is
therefore not reported.
```

```
Information criteria:
Akaike:            -1.11718E+01
Hannan-Quinn:      -1.11712E+01
Schwarz:           -1.11681E+01
Model type: OLS with GARCH errors

Error specification:
[1 - r(1,1)L]V(t) = d + [r(2,1)L]u(t)^2,
where u(t) is the error of the OLS conditional expectation model,
V(t) is its conditional variance,
d is a constant, and L is the lag operator.
All the parameters should be non-negative.
However, only the restrictions V(t) > 0 will be imposed.

Model type: OLS with GARCH errors

Error specification:
[1 - r(1,1)L]V(t) = d + [r(2,1)L]u(t)^2,
where u(t) is the error of the OLS conditional expectation model,
V(t) is its conditional variance,
d is a constant, and L is the lag operator.
All the parameters should be non-negative.
However, only the restrictions V(t) > 0 will be imposed.

The likelihood has been maximized using the simplex method of
Nelder and Mead.
The algorithm involved is a Visual Basic translation of the
Fortran algorithm involved in:
W.H.Press, B.P.Flannery, S.A.Teukolsky and W.T.Vetterling,
'Numerical Recipes', Cambridge University Press, 1986,
pp. 292-293

Maximum likelihood estimation results:

Parameters  ML estimate             t-value [p-value]
b(1)         -0.000094                     -1.925 [0.05423]
d              0.000000                          7.124[0.00000]
r(1,1)        0.894137                   239.688 [0.00000]
r(2,1)        0.072464                     20.950 [0.00000]
[The two-sided p-values are based on the normal approximation]

RSS = 54.803603120E-03
s.e. = 37.519958939E-04
R-square = -0.0006
n = 3897

If the model is correctly specified, in the sense that the conditional
expectation of the error e(t) relative to the X variables and all
lagged dependent (Y) variables and lagged X variables equals zero, then
the OLS-GARCH parameter estimators minus their true values, times the
square root of the sample size n, are (asymptotically) jointly normally
distributed with zero mean vector and variance matrix :
```

```
 9.37509912E-06 -4.23373768E-09 3.55517980E-04 -2.17479503E-04
-4.23373768E-09 1.65900830E-11 -5.98748517E-07 -1.46021836E-07
 3.55517980E-04 -5.98748517E-07 5.42306258E-02 -2.78703020E-02
-2.17479503E-04 -1.46021836E-07 -2.78703020E-02 4.66241587E-02
```

Le coefficient $r(1, 1)$ correspond à β_1, le coefficient d à α_0 et le coefficient $r(2, 1)$ à α_1. Les valeurs obtenues pour les coefficients estimés sont un peu différentes des valeurs estimées obtenues avec TSP. Même si vous utilisez des valeurs de départ égales aux valeurs estimées obtenues avec TSP, cela ne change rien aux résultats obtenus avec Easyreg. La raison de cette disparité : les deux logiciels utilisent des algorithmes d'optimisation différents. Les résultats sont toutefois qualitativement similaires et l'ordre de grandeur des coefficients estimés est très proche.

Références bibliographiques

[BOL 1992] T. Bollerslev, R.S. Chou, K.F. Kroner, ARCH modelling in finance – A review of the theory and empirical evidence, *Journal of Econometrics*, 52, p. 5–59, 1992.

[MCF 1976] McFadden, Quantal Choice Analysis, A Survey, *Annals of Economic and Social Measurement*, 5, 363–390, 1976.

[PED 2003] C. Pederzoli, Stochastic volatility and GARCH a based on UK stock data, working paper, University of Modena and Reggio Emilia, 2003.

[TAY 1986] S.J. Taylor, *Modelling financial time series*, John Wiley and Sons, Chichester, 1986.

[TAY 1994] S.J. Taylor, Modeling stochastic volatility: a review and comparative study, *Mathematical Finance*, 4, p. 183–204, 1994.

Index

V

Sciences de gestion



Économétrie

L'auteur :

Éric Dor, directeur de la recherche et professeur associé d'économétrie à l'IESEG School of Management (Lille), enseigne également à l'Institut Catholique des Hautes Études Commerciales (ICHEC) de Bruxelles.

Cet ouvrage présente en sept chapitres les méthodes essentielles de l'économétrie et explique comment les appliquer à des problèmes concrets au moyen d'outils informatiques appropriés. Tout en gardant une grande rigueur mathématique et statistique, il expose de façon claire et pédagogique toutes les notions importantes.

Le livre comprend une grande variété d'exercices d'application, qui portent aussi bien sur des problématiques d'entreprise que des problématiques financières et macroéconomiques. La solution est systématiquement détaillée, de la spécification du modèle à son estimation et à l'interprétation des résultats. Le lecteur est ainsi initié à la résolution de problèmes économétriques au moyen des logiciels Excel, SPSS, TSP et Easyreg.

L'ouvrage s'adresse aux étudiants d'économie et gestion en licence et master ainsi qu'aux étudiants en écoles de commerce. Il constituera aussi un outil précieux pour les praticiens.

Les données des exercices, au format Excel, sont disponibles sur le site www.pearson.fr.

Direction de collection-:

Roland Gillet, professeur à l'université Paris 1 Panthéon-Sorbonne

Dans la même collection :

- **Finance**, André Farber *et al.* (Solvay Business School, ULB, Bruxelles)
- **Probabilités, statistique et processus stochastiques**, Patrick Roger (ULP, Strasbourg)
- **Mathématiques appliquées à la gestion**, Ariane Szafarz *et al.* (Solvay Business School, ULB, Bruxelles)

La collection Synthex propose aux gestionnaires et aux économistes de découvrir ou de réviser une discipline et de se familiariser avec ses outils au travers d'exercices résolus.

Chaque ouvrage présente une synthèse pédagogique et rigoureuse des techniques et fondements théoriques, qu'une sélection d'exercices aux corrigés détaillés permet d'assimiler progressivement. Le lecteur, étudiant ou professionnel, est ainsi conduit au cœur de la discipline considérée, et, via la résolution de nombreux problèmes, acquiert une compréhension rapide et un raisonnement solide.

ISBN : 978-2-7440-4071-9

PEARSON

Pearson Education France
47 bis, rue des Vinaigriers 75010 Paris
Tél. : 01 72 74 90 00
Fax : 01 42 05 22 17
www.pearson.fr

